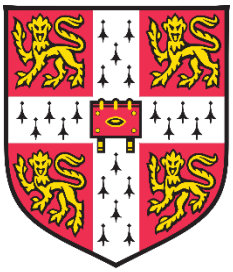
GLUCURONOXYLAN ACETYLATION PATTERNS AND SITE-SPECIFIC ACETYLTRANSFERASES

by

JOEL WURMAN RODRICH



University of Cambridge

Magdalene College

This dissertation is submitted for the degree of

Doctor of Philosophy

February 2019

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Summary

Glucuronoxylan Acetylation Patterns and Site-Specific Acetyltransferases

Joel Wurman-Rodrich

Plant biomass is the biggest source of carbon-based materials for several industries such as food, agriculture, textile, construction, and biofuels, among others. Cellulose and xylan are the two major components of plant biomass, and their intermolecular interactions in the cell wall are crucial for plant growth and industrial use. In eudicots, these interactions are determined by the arrangement of the main substitutions on the xylan backbone, *O*-acetyl (Ac) and glucuronic acid (GlcA) groups. GUX1, GUX2 and GUX3 are largely responsible for transfer of GlcA decorations, each producing a distinct substitution pattern in regions of the xylan molecule known as compatible or incompatible domains, based on the likely ability of the xylan produced to bind to hydrophilic surfaces of cellulose fibrils. However, the elongation of the xylan backbone, the origin of the GlcA patterning, and the xylan acetylation processes are not fully understood.

The xylan backbone is synthesised by two partially redundant Xylan Synthase protein Complexes (XSC). Removing subunits of one XSC causes a reduction in xylan length, and affects the GlcA pattern. Xylan acetylation has also been shown to modulate the GlcA pattern, thus, changes in XSC could also affect the acetylation. In the work described in this thesis, mutants with altered XSC composition and no GlcA on xylan were studied. They showed no detectable changes in xylan acetylation patterns. Therefore, the effect on GlcA patterning can be attributed to the shortening of the xylan backbone.

Mainly, every other xylose residue is acetylated, but the pattern in other parts of the molecule is unclear. The xylan compatible and incompatible domains were analysed using a combination of xylan-specific hydrolases, carbohydrate gel electrophoresis and tandem mass-spectrometry. The structures found in the compatible domain were consistent with the reported acetylation pattern, but the incompatible domain exhibited consecutive acetylated xylosyl residues. *Arabidopsis* mutants for candidate xylan *O*-acetyltransferases were analysed, and four genes were found to be involved in xylan acetylation. TBL32 and TBL33 were shown to be responsible for the acetylation of xylosyl residues that carry a GlcA. Mutants in TBL3 and TBL28 also affected xylan acetylation, opening opportunities for discovering new xylan acetylation functions.

Para mi mamá y mi papá

(For my mum and my dad)

... שֶׁהֶחֱיָנוּ וְקִיְּמָנוּ וְהִגִּיעָנוּ לַזְּמַן הַזֶּה.

(…who has granted us life, sustained us and enabled us to reach this occasion.)

Acknowledgements

First of all, I want to thank my tutor and mentor Prof Paul Dupree for the enormous opportunity that he has given me, by inviting me to join his lab in 2013, and for the wonderful five and a half years of great scientific collaboration.

I want to thank my mum Jenny Rodrich Eskenazi and my dad Leon Wurman Shapiro, for their continuous unconditional love and support. I have been able to reach this goal thanks to both of them, and this is why this thesis is dedicated to them. I want to thank my brothers Dani and Alan for believing in me and for encouraging me during this period of my life, and my sisters Fran, Einat and Sari who also have been very supportive.

I want to thank my former tutor Prof Mike Handford from Universidad de Chile for opening up new doors in my career and for all the knowledge that he transmitted to me. I want to thank Prof Maija Tenkanen, Dr Sun-Li Chong, Prof Ykä Helariutta and Dr Juha Immanen from the University of Helsinki for the great collaboration on the birch xylan project, and for kindly providing the birch material. Thanks to Prof Yihua Zhou and Dr Baocai Zhang from the Institute of Genetics and Developmental Biology in Beijing, for the collaboration on ESK1 and TBL33 projects, Prof Kris Krogh from Novozymes for the GH30 project collaboration.

I want to thank my fabulous friends Susan, Johana, Aleix, Greg, Jeff, Chloe, Alex, Matthieu, Hector, Sergio and Damian, who have made this time of my life in Cambridge sparking and joyful. I want to thank my friends who have sent me their love from the distance; Moti, Raimundo, Ariel Stern, Ariel Cerda, Talo, Mati and Aitor. Thank you all so much for all the love.

I want to thank all my wonderful lab mates: Fede, Dora, Nino, Tias, Nadine, Matteus, Nick, An, Nuno, Carolina, Rita, Marta, Tom, Denis, Xiaolan and Wendy who kindly helped me to adapt to the new environment; also Henry, Li, Oliver, Jan, Ian, Louis, Yoshi, Prof Toshi Kotake, Clelton, Lara, Paolo, Jack and Chris G. It has been a real pleasure for me to work in such a fun and interactive atmosphere that I have got daily from all of you.

Finally yet very importantly, I want to thank my partner in crime and exquisite lover. Thank you Dean for every amazing moment we have lived together so far, and all the humongous amount of love and support you have given me every day.

x

List of Abbreviations

[Me]GlcA	Glucuronic acid or 4-*O*-Methylglucuronic acid residue
°C	Celsius degrees
2AA	2-Aminobenzoic acid
2-PB	2-picoline-borane
Ac	*O*-acetyl group
AcN	Acetonitrile
AIR	Alcohol-insoluble residue
AmAc	Ammonium acetate ($NH_4CH_3CO_2$)
ANTS	8-aminonaphthalene-1,3,6-trisulfonic acid
APTS	8-aminopyrene-1,3,6-trisulfonic acid
Ara*f*	Arabinofuranosyl residue
Ara*p*	Arabinopyranosyl residue
bp	Base pairs
CAZy	Carbohydrate-active enzymes
cDNA	complementary DNA
D_2O	Deuterium oxide
DA	Degree of acetylation
DASH	DNA sequencer-assisted saccharide analysis in high throughput
DHB	Dihydroxybenzoic acid
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

DP	Degree of polymerisation
EDTA	Ethylenediaminetetraacetic acid
Gal	Galactosyl residue
GalA	Galacturonic acid residue
GH	Glycosyl hydrolase
GlcA	Glucuronic acid residue
GT	Glycosyltransferase
GUX	GLUCURONIC ACID SUBSTITUTION OF XYLAN
GXM	Glucuronoxylan methyltransferase
HCl	Hydrochloric acid
HG	Homogalacturonan
HILIC	Hydrophilic-interaction liquid chromatography
IRX	IRREGULAR XYLEM
KO	Knock-out
Kb	Kilo-base pairs
MeGlcA	4-*O*-Methylglucuronic acid residue
ML	Maximum likelihood
MS	Mass spectrometry
N	Pentosylated Xyl
$NaBH_4$	Sodium borohydride
NaOH	Sodium hydroxide
NMR	Nuclear magnetic resonance
NRE	Non-reducing end

OD_{600}	Optical density measured at a wavelength of 600 nm
PACE	Polysaccharide analysis using carbohydrate gel electrophoresis
PCR	Polymerase chain reaction
Pent	Pentosyl residue
R	Mono or di-substituted Xyl resistant to acetylesterase CE4
RCF	Relative centrifugal force
RE	Reducing end
REO	Reducing-end oligosaccharide
RG	Rhamnogalacturonan
Rha	Rhamnosyl residue
RNA	Ribonucleic acid
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
RWA	REDUCED WALL ACETYLATION
SDS	Sodium dodecyl sulfate
T-DNA	*Agrobacterium*-mediated transfer DNA
TAIR	The Arabidopsis information resource
TBL	TRICHOME BIREFRINGENCE-LIKE
UDP-GlcA	Uridine 5’-diphosphate glucuronic Acid
UDP-Xyl	Uridine 5’-diphosphate xylose
WT	Wild-type
Xyl	Xylosyl residue

Table of Contents

Table of Figures

List of Tables

Chapter 1: Introduction

1.1 Context

Plant-based materials such as wood and its derivatives have been used by society since before the beginning of human history. In 2016, more than 5500 million m^3 of wood and derivatives were produced in the world (Food and Agriculture Organization - FAO, 2018). Also, land plants are an almost carbon-neutral resource as they fix atmospheric CO_2, and compose the largest biomass on earth (Bar-On et al., 2018). It has been estimated that about 70% of the net CO_2 fixation in land plants is deposited as newly formed plant cell wall material, approximately 150-170 billion tons per year (Pauly and Keegstra, 2008).

It is the molecular architecture of the wall itself which provides a challenge for the use of this material in industrial applications (McCann and Carpita, 2015). Therefore, to develop more efficiently extractable materials, and to establish new extraction strategies that would allow a better exploitation of the lignocellulosic biomass, it is very important to understand how the cell wall constituents interact.

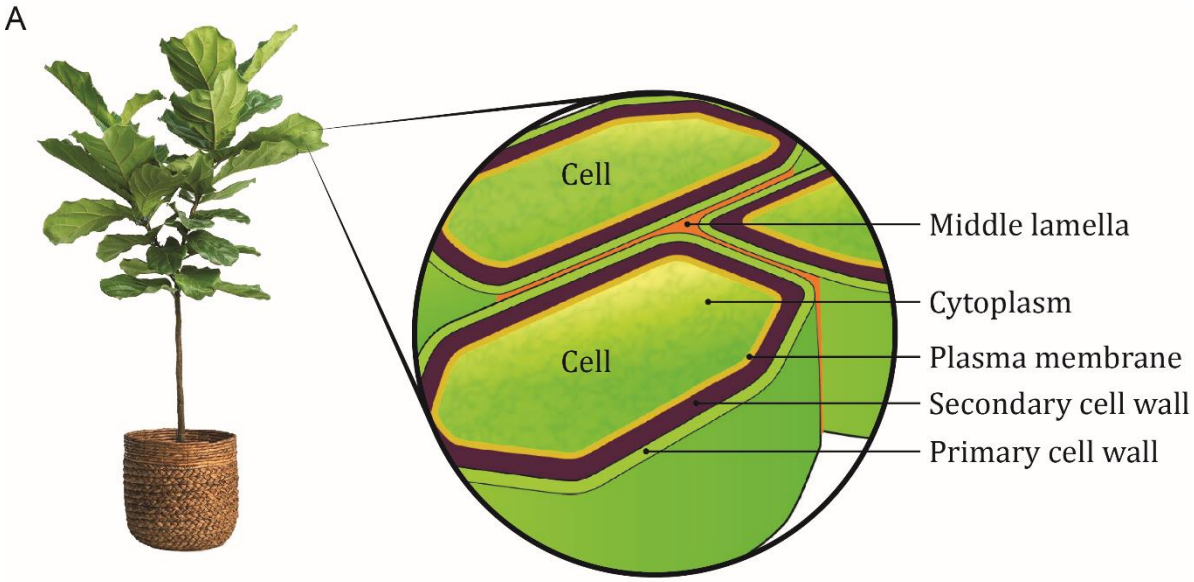
1.2 Components and organisation of the plant cell wall

Land plants originated in the Middle Ordovician period, around 470 million years ago (Rubinstein et al., 2010; Wellman et al., 2003). It is believed that land plants emerged from charophytes, a green algal family, of which some members developed cellulosic biomass in their cell walls. The change in habitat, from water to land pressured plants to develop several characteristics that enabled them to adapt into this new environment. Possibly, cellulose provided enough strength to the first land plants to support their own weight in order to gain height and compete for the light. Some tree canopies can have over 40 meters of diameter (Figure 1.1), and therefore, the cell walls in their branches are required to provide support for resisting immense mechanical stress.

Figure 1.1: The Hitachi tree.
The canopy of this Monkeypod tree in Moanalua gardens in Hawaii, USA, has approximately 40 metres of diameter. Photo by Bhawna Mundotia (www.hitachi.com/hitachitree/en/).

Cell walls coat every single plant cell on earth, determining their strength and elasticity, and granting other physical properties that allow plants to live and stand on land. The cell wall is mainly composed by cellulose, hemicelluloses and pectin (Figure 1.2), and the proportions of these components and their interactions in the wall determine properties such as permeability and connectivity to other cells. Plant cell walls are also durable barriers against certain environmental stresses and pathogenic organisms. However, how the wall composition determines all of these properties is poorly understood. The work carried out for this thesis was oriented towards investigating the molecular structure of the main hemicellulose of the wall; the xylan.

A



B

Figure 1.2: Primary and secondary cell wall models.
Panel A: cellular localisation of plant cell walls. Panel B: diagrams of primary and secondary cell walls. Graphics on top panel are modified from two websites: science.energy.gov/ber and greenerynyc.com. Diagrams at the bottom are modified from Loix et al (2017).

Cells firstly produce a primary cell wall in which they proliferate and elongate. The primary wall is constituted by long and crystalline cellulose microfibrils embedded in a hydrated matrix of polysaccharides and glycoproteins (Cosgrove, 2005). In the majority of higher plants (including *Arabidopsis*) this non-cellulosic matrix is formed mostly of xyloglucan and pectin. The interactions between cellulose and xyloglucan are believed to control the process of cell expansion (Cosgrove, 2014). Small amounts of xylan have been described in primary cell walls (Darvill et al., 1980; Mortimer et al., 2015; Zablackis et al., 1995), but the function of xylan in primary cell walls is not understood. A typical example of primary cell wall is found in liquid-grown callus. Callus cell walls are composed of abundant polygalacturonan, rhamnogalacturonan I and xyloglucan, with minor amounts of type-II arabinogalactan, extensins, rhamnogalacturonan II, glucomannan and glucuronoarabinoxylan (Barton et al., 2006; Goubet et al., 2002; Handford et al., 2003; Manfield et al., 2004). Consistent with this composition, several polysaccharide synthesis enzymes are expressed and active in callus cells, including enzymes necessary for the synthesis of xylan (Nikolovski et al., 2012; Sorieul et al., unpublished).

The termination of the cell growth is sometimes followed by the development of a secondary cell wall, which grants further fortification and strength (Scheller and Ulvskov, 2010). This is where a crucial function of xylan takes place. The secondary cell wall comprises of a “reinforced concrete”-like structure made of cellulose microfibrils cross-linked by xylan and other non-cellulosic polysaccharides, embedded in a matrix of lignin and proteins (Fry, 1986). This elaborate organisation of the wall is critical for the maintenance of the cell structure of fibres and xylem vessels, and confers mechanical support for the plant (Fernandes et al., 2011; Grantham et al., 2017; Simmons et al., 2016). To introduce the xylan context in the cell, other components of the wall are next briefly presented and schematised.

1.2.1 Cellulose

Cellulose is the most abundant polymer in both primary and secondary walls and it grants most of the supportive strength that the walls need (Somerville, 2006). Cellulose chains are composed of β-1,4-linked D-glucosyl residues without any branches. The minimal repeating unit of the cellulose structure is cellobiose with a β-

bond rotation of 180°. The extent of these chains is unclear but it is believed that they can reach thousands of residues long (Cosgrove, 2014). Cellulose chains bundle into paracrystalline superstructures, named microfibrils (Figure 1.3). Two forms of cellulose crystals have been described in plants (two-chain monoclinic I*β* and single-chain triclinic I*α*), which differ in the organisation of the cellulose chains but form part of the same microfibrils and (Atalla and Nagel, 1984; Nishiyama et al., 2003, 2002). Cellulose microfibrils are located parallel to the plasma membrane, and in a single microfibril the chains are parallel to each other. The microfibril sectional area and shape are unclear, but these geometrical features are important determinants of the physical properties of cellulose and its interactions with the matrix components (Cosgrove, 2014). Hexagonal and square fibril sections have been proposed, each presenting different hydrophobic and hydrophilic faces for interactions determined by these two physical properties (Cosgrove, 2014). It has been estimated that between 18 and 36 chains conform a single cellulose microfibril, but the quantity varies between plant species and the measuring methods (Hill et al., 2014; Jarvis, 2013; Newman et al., 2013; Nixon et al., 2016).

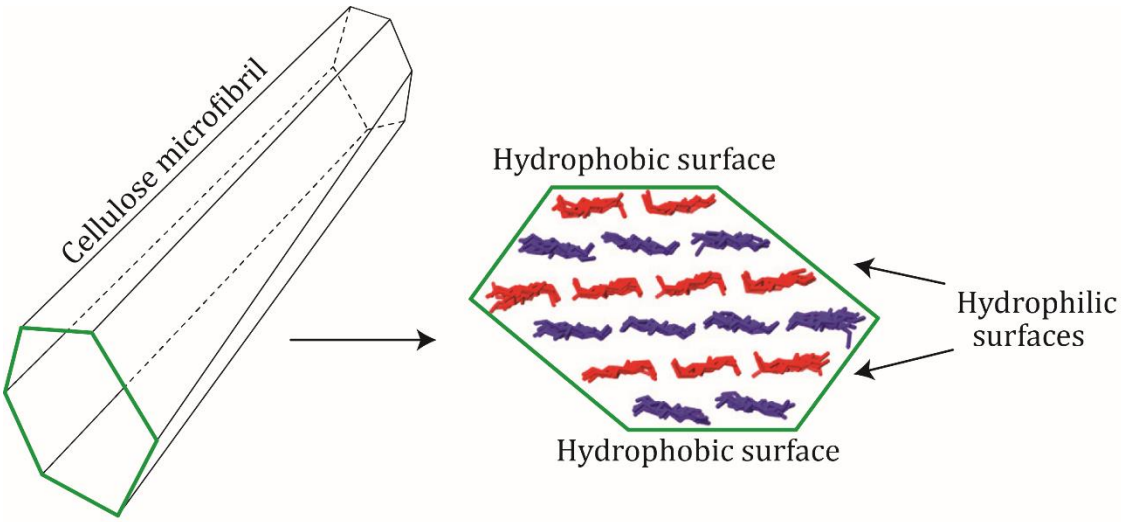


Figure 1.3 Section of a hexagonal 18-chain conformation of a microfibril.
The section view is from the non-reducing end after a molecular-dynamics simulation for 200 ns. The top and bottom surfaces of the section have hydrophobic characteristics while the other four surfaces are hydrophilic. Only one of many possible conformations of cellulose microfibrils is shown. Modified from Oehme et al. (2015).

The cellulose microfibrils are synthesised by the cellulose synthase complex (CSC) at the plasma membrane (Kimura, 1999; Somerville, 2006). The CSC is composed by members of the CESA protein family which form six-membered multimers termed

rosettes, and the number of subunits per multimer is still under debate (Figure 1.4). A recent model has been proposed in which each multimer of the rosette (hexamer) would contain three active CESAs, and produce microfibrils with 18 cellulose chains (Gonneau et al., 2014; Hill et al., 2014; Jarvis, 2013). Each CESA in the complex is thought to catalyse the synthesis of a single cellulose chain (Hill et al., 2014; Somerville, 2006). The composition of the rosette varies between primary and secondary cell wall CSCs. CESA1, 3, and 6-like (CESA6, 2, 5, and 9) proteins collaborate in the synthesis of cellulose in primary walls, while CESA4, 7, and 8 are each required for cellulose biosynthesis in secondary cell walls (Desprez et al., 2007; Persson et al., 2007; Taylor et al., 2003). During secondary cell wall synthesis, dense clusters of CSCs move together in a single direction, synthesizing highly aggregated cellulose microfibrils. In contrast, in primary cell wall synthesis, several isolated CSCs move along shared tracks in different directions while synthesizing lowly aggregated cellulose microfibrils (Li et al., 2014). The high specialisation of both sets of CSC and the difference in the mechanism in which the cellulose is made in both primary and secondary cell walls suggests that the function of cellulose, or its interactions with xylan and other components might be essentially different.

Figure 1.4: Hypothetical scheme of cellulose microfibril synthesis.
Cellulose synthesis by the CesA is illustrated. 18 cellulose chains (blue and light blue) emerging from its constituent CesA catalytic subunits. The chains bend through 90° and crystallize into a microfibril that lies against the inner face of the cell wall. Three CESAs may interact to form a trimeric particle, which in turn may assemble into a hexameric rosette. Only the transmembrane domains of the 18 CESAs are shown: there is considerable expansion of the structure inside the cell membrane, not shown here. Modified from Jarvis et al (2013).

1.2.2 Hemicelluloses

Hemicelluloses are a group of polysaccharides of the cell wall, which are not cellulose, and predominantly have a β-1,4-linked backbones of pyranosyl residues (glucose, mannose or xylose) (Scheller and Ulvskov, 2010). Hemicelluloses have been described as polymers from the cell wall that are solubilised by alkali solvents, and are β-1,4-linked pyranosyl residues that have the same equatorial configuration at C1 and C4. Therefore, the backbones of these glycans have a significant structural similarity (Figure 1.5). These characteristics allow them to adopt a cellulose-like conformation and form hydrogen bonds to cellulose chains (Caffall and Mohnen, 2009). Xylan, xyloglucan, mannans fit with this technical definition but mixed-linked glucan and type-II arabinogalactans are also considered hemicelluloses (Caffall and Mohnen, 2009; Fry et al., 2008).

Figure 1.5: Hemicellulose repeating disaccharides.
β-1,4 glycosidic bonds between hemicellulose backbone residues have equatorial configurations at C-1 and C-4. Diagram on top shows the numbering of sugar carbon positions. White – glucosyl residue, Purple – Xyl, Green – mannosyl residue. Modified from (Scheller and Ulvskov, 2010).

1.2.2.1 Xylan

Xylan is the most abundant hemicellulose on earth and its role is vital for plants (Dodd and Cann, 2009; Hao and Mohnen, 2014). Xylan is composed by a linear β-1,4-D-Xyl*p* chain and depending on the sidechain substitutions it is classified into four different types depending on the backbone modifications: arabinoxylan, glucuronoarabinoxylan, glucuronoxylan, or unsubstituted homoxylan. The xylan backbone Xyl can also be found modified with *O*-acetyl substitutions (Ac) at O-2 and/or O-3 (Chesson et al., 1983). In eudicot secondary cell walls, the xylan backbone is heavily acetylated (Busse-Wicher et al., 2014; Evtuguin et al., 2003; Gonçalves et al., 2008; Teleman et al., 2002, 2000), however, little is known about the function of Ac on xylan. The work of this thesis was focused on describing the arrangement of Ac on the xylan backbone to investigate the possible roles that this modification may have. A more detailed description of the xylan structure is presented in a dedicated section in this chapter (section 1.3, page 14). Some of the other cell wall hemicelluloses are also acetylated, and their structures are next briefly described.

1.2.2.2 Xyloglucan

Xyloglucan is in every land plant that has been analysed, and its function is related to cell elongation and carbon storage (Pauly et al., 2013; York et al., 1984). Xyloglucan is the most abundant hemicellulose in primary cell walls of angiosperms and gymnosperms, except for grasses (Scheller and Ulvskov, 2010). In eudicots, xyloglucan constitutes around 20–30% of primary cell walls (Fry and Miller, 1989). It has a repeating β-1,4 glucan backbone decorated with a regular pattern of α-1,6-linked Xyl (Figure 1.6). The Xyl branches can be further substituted with different sugars at O-2 (Pauly et al., 2013), and β-1,2-Gal (galactosyl residue) on Xyl can carry α-1,2-fucosyl side chains (Ivakov and Persson, 2012; Kiefer et al., 1989). Xyloglucan is an acetylated hemicellulose, and in *Arabidopsis* it has been reported to be exclusively acetylated on the Gal (Kiefer et al., 1989). Ac are mainly linked to O-6 but have also been found at O-3 and O-4, and in WT about 75% of these sites are acetylated (Gille et al., 2011; Kiefer et al., 1989). Xyloglucan is covalently linked to rhamnogalacturonan I (RG-I, section 1.2.4), and has been proposed to function as cross-linker between

cellulose microfibrils, forming a strong xyloglucan-cellulose structural network in primary cell walls (Pauly et al., 2013; Popper and Fry, 2008).

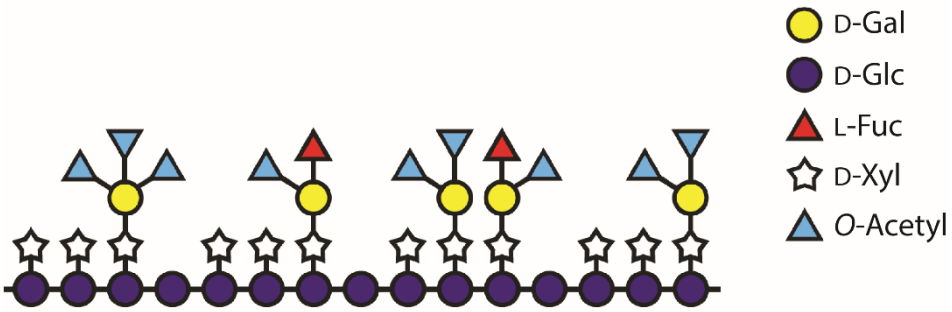


Figure 1.6: Xyloglucan structure.
Modified from Burton et al (2010).

1.2.2.3 Mixed-linkage glucan

Mixed-linkage glucans (MLG) are considered hemicelluloses, despite the fact that the backbone is non-linear. They are most commonly found in grasses, but also in early-diverging plant lineages, like *Equisetum* (Buckeridge et al., 2004; Fry et al., 2008; Sørensen et al., 2008). MLG is an unbranched homopolymer with no backbone modifications, and is not acetylated (Figure 1.7). The backbone is constituted by stochastically distributed β-1,4-linked cellotriose and cellotetraose units connected by β-1,3 linkages (Buckeridge et al., 2004). The combination of these two linkages make the backbone linear but kinked, which possibly grants flexibility and solubility to the polymer (Buliga et al., 1986). The role of MLG has been associated with storage, cell expansion, and also in some cases to provide structural support. It has been proposed that MLG may also play a role in growth termination and transition to secondary cell wall production (Vega-Sanchez et al., 2012).

Figure 1.7: Mix-linked glucan structure.
Mixed-linked glucan is a linear structure of β-1,4-cellotriose and β-1,4 cellotetraose
Modified from Burton et al. (2010).

1.2.2.4 Mannans

Mannans are the predominant hemicellulose in secondary cell walls of gymnosperms, but are also present in angiosperms (Pauly and Keegstra, 2008). The mannans are classified into four different groups according to their backbone and sidechains composition: mannan, glucomannan, galactomannan and galactoglucomannan. While mannan and galactomannan consist of a backbone made of only mannose, glucomannan and galactoglucomannan possess both mannose and glucose backbone units (Figure 1.8) (Pauly et al., 2013). The mannosyl residues in galactomannan and galactoglucomannan can carry α-1,6-Gal (Scheller and Ulvskov, 2010). Some mannans are acetylated; backbone mannosyl residues can carry one or two Ac, at O-2 and/or O-3 positions (Manna and McAnalley, 1993; Teleman et al., 2000).

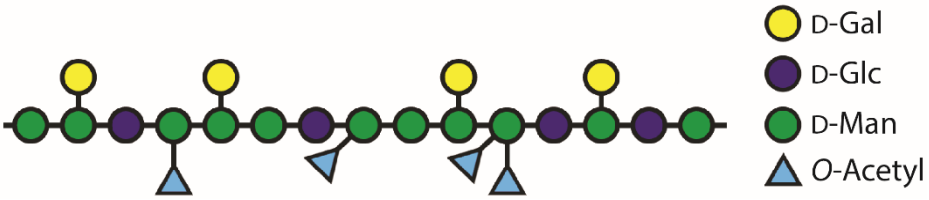


Figure 1.8: Galactoglucomannan structure.
Modified from Burton et al. (2010).

1.2.2.5 Arabinogalactans

Arabinogalactans (AGs) are structurally complex largely-branched polysaccharides, often attached to hydroxyproline residues of several plant cell wall proteins (AGPs) (Tryfona et al., 2012). AG chains in AGPs can have 150 or more sugar residues (Kieliszewski, 2001; Showalter, 2001). Their structure is defined by a backbone of β-1,3-D-Gal*p*, and sidechains of β-1,6-D-Gal*p* decorated with L-Ara*f* and occasionally with other sugars like D-Gal*p*, L-Ara*f*, L-Rha*p* (rhamnopyranose), and D-GlcA (Carpita and Gibeaut, 1993; Hinz et al., 2005). There are two types of AG described; AG-II which may also be substituted with a non-reducing-end β-1,6-GlcA decoration, is far more widespread and is usually found as the glycosidic parts of AGPs (Kotake et al., 2006; Tryfona et al., 2012, 2010); AG-I have a different galactan backbone structure that is connected to RG-I (section 1.2.4), and is considered a side-chain with an extension of up to 50 glycosyl residues long (Nakamura et al., 2002).

1.2.3 Lignin

Lignin is the third mayor component of the cell wall after cellulose and hemicelluloses. It is a complex polyphenolic macromolecule highly abundant in secondary cell walls, accounting about 18–35% of the plant's dry weight (Pauly and Keegstra, 2008). It is also an essential element for water conduction and defence systems of plants, and it contributes considerably to the compressive strength of secondary xylem tissues. The lignin polymer is made by the free-radical polymerisation of three monomers or monolignols which differ in the methoxylation degree of the aromatic ring; *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (Figure 1.9) (Boerjan et al., 2003). The three monolignols are synthesised in the cytoplasm, and then they are exported into the wall, where laccases and peroxidases activate their polymerisation (Ros-Barceló et al., 2006). Lignin is embedded in the wall of plant vasculature cells, particularly in the xylem vessels elements and fibres that constitute most of secondary xylem tissues (Mottiar et al., 2016). Lignin is also present in the sclerenchyma fibres and sclereids in xylem and phloem, and in the cortex cells of the periderm. The hydrophobic nature of lignin confers water impermeability to the cell wall of these cells, allowing the vessel cells to perform as pipes for water transport (Bonawitz and Chapple, 2010).

The lignin polymer is acetylated. Acetyl groups are found linked to the γ-carbon of S and G monomers (Del Río et al., 2007). In hardwood xylem, between 1 and 50% of the lignin available sites can be acetylated, while in softwoods xylem acetylation has not been reported.

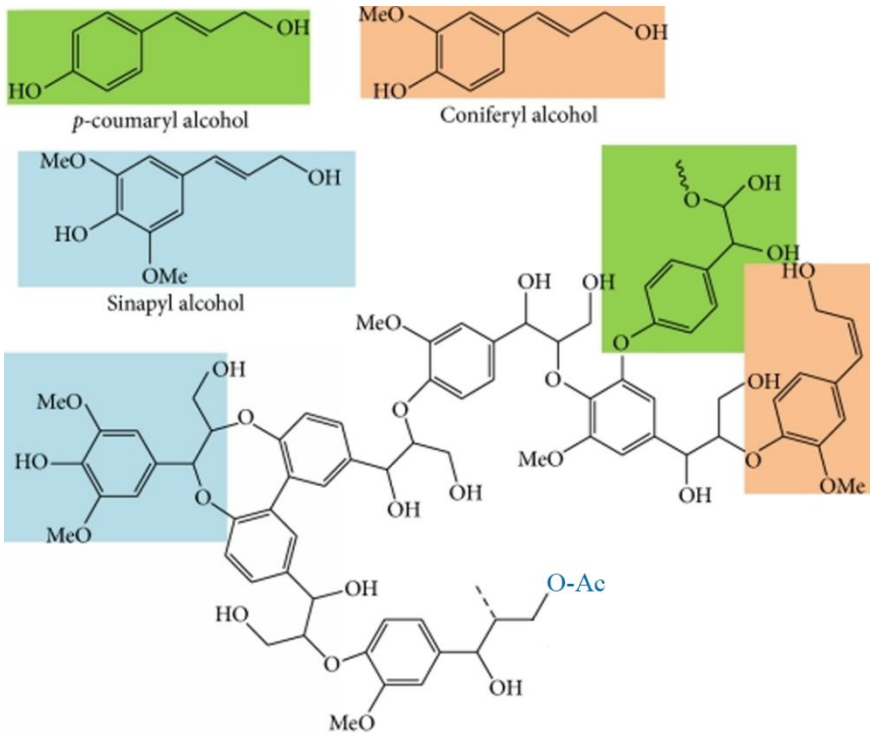


Figure 1.9: Lignin precursors and structure.
p-Coumaryl, Sinapyl and Coniferyl alcohols are shown as substrate and residues integrated into the lignin structure. An *O*-acetyl group is shown in blue. Modified from Lee et al. (2014).

1.2.4 Pectins

Pectin is one of the most complex polysaccharides on earth, and is present in charophytes and all land plants (O'Neill et al., 2004). Pectins are abundant in walls of growing and dividing cells, walls of cells in soft parts of the plants, in the middle lamella and the cell corners (Mohnen, 2008). Around 35% of dicots primary cell walls are made of pectin, and about 5% of woody tissue as well (O'Neill et al., 1990). The main pectic polysaccharides are four; Homogalacturonan (HG), rhamnogalacturonan I (RG-I), rhamnogalacturonan II (RG-II) and xylogalacturonan (see Figure 1.10). HG is a polymer of α-1,4-linked galacturonic acid (GalA) that embodies 60% of pectins, and in *Arabidopsis* 23% of leaf walls is HG (Mohnen, 2008). RG-II is formed when clusters of complex sidechains are O-2 and O-3-linked to a GalA backbone. These side-chains are composed of 12 different types of glycosyl residues linked together with at least 20 different glycosidic bonds and boron (Matsunaga et al., 2004; O'Neill et al., 2004). Xylogalacturonan has a homogalacturonan backbone, and 25-75% of backbone GalA have a β-1,3-Xyl side-chain (Le Goff et al., 2001; Schols et al., 1995). RG-I is the only

pectin without a pure GalA backbone. Instead, it is a branched polymer with a backbone of α-1,4-D-GalA-α-1,2-Rha disaccharide repeats, where Rha can be substituted with β-1,4-galactan, branched arabinan and/or arabinogalactan sidechains. (Harholt et al., 2010). Pectins are successive polysaccharide domains that are covalently bound into single molecules (Harholt et al., 2010). Arabinan and Type-I arabinogalactan are covalently linked to RG-I, while RG-II is thought to be linked to HG (Coenen et al., 2007). Also, a link between pectins and hemicelluloses have been proposed by the recent description of a molecular structure in primary cell walls. Arabinoxylan Pectin Arabinogalactan Protein 1 (APAP1) is a type-II AGP structure that was shown to have a covalent link to arabinoxylan and RG-I (Tan et al., 2013).

Pectin polysaccharides are acetylated and methylated (Ishii, 1997, 1995; Komalavilas and Mort, 1989). Pectin acetylation and methylation has been proposed to interfere with in the formation of Ca^{+2}-mediated interactions between pectin chains (Liners et al., 1992). Methyl-esterification occurs at C-6 of GalA. In RG-I, acetylation has been reported only at O-3 of GalA, but GalA in pure galacturonan backbones can be acetylated at O-2 and/or O-3. The position of the Ac on pectin backbone GalA are analogue to the positions where Ac can be found on xylan backbone Xyl. Therefore, similarities between the acetylation processes of both polysaccharides are possible.

Figure 1.10: Pectin domains diagram.
Schematic representation of the four main pectic domains and their common side-chain types. Abbreviated sugars: Dha – 3-deoxy-D-*lyxo*-2-heptulosaric acid, Kdo – 2-keto-3-deoxy-D-*manno*-octulosonic acid. Modified from Scheller et al. (2007).

1.3 Xylan backbone: structure and biosynthesis

Xylan is the most abundant hemicellulose constituting up to 30% (w/w) of biomass dry-weight in some hardwood species (Busse-Wicher et al., 2016b). Xylans are present in both primary and secondary cell walls of higher plants. In eudicots, xylan is predominantly in secondary cell walls, while high amounts can be found in monocot primary and secondary cell walls. Contrastingly, in gymnosperms, xylan is a relatively minor component of the cell wall, and its role is unclear (Pauly and Keegstra, 2008).

1.3.1 Backbone structure

The backbone of xylan is composed by xylopyranosyl residues linked by β-1,4 glycosidic bonds. Xylose, like many other monosaccharides, is a reducing sugar (aldose) as it has an aldehyde group at the anomeric carbon (C-1) (Nelson and Cox, 2013). Aldoses in solution are in equilibrium between open form (acyclic) and a more predominant closed form (cyclic) (Figure 1.11). When in open form, xylose exposes the aldehyde group, which is a reducing agent. When two or more xyloses are 1,4-linked, the C-1 of one monomer is linked to the C-4 of the next. The Xyl with a free C-1 can switch between open and closed states, therefore this Xyl is at the ‘reducing end’ (RE) of the chain. Xyl without a free C-1 are called non-reducing, as they are permanently forming a ring (Nelson and Cox, 2013).

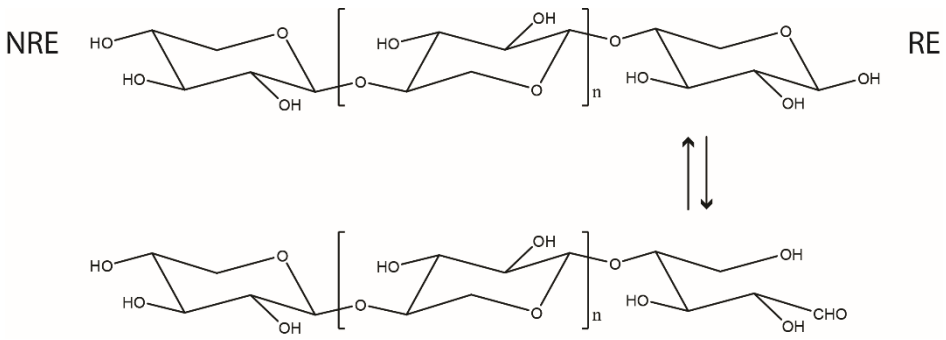


Figure 1.11: Open and closed forms equilibrium of RE Xyl.
The top structure shows the cyclic form (closed) of the xylosyl residue at reducing-end (RE) of a xylooligosaccharide. The bottom structure shows the less predominant acyclic form (open). The structure spontaneously swaps between forms with a predominance for the cyclic form. Other xylosyl residues of the chain are permanently on a cyclic form, and the terminus with a non-reducing Xyl is denominated non-reducing end (NRE).

1.3.2 Initiation of synthesis

It is unclear how xylan synthesis is initiated, however, a specific sequence of sugar residues at the RE of xylan (Peña et al., 2007) suggests that this sequence acts as a primer for xylan synthesis. (Figure 1.12). The oligosaccharide sequence at the RE of xylan (REO) is 4-β-D-Xyl-1,4-β-D-Xyl-1,3-α-L-Rha-1,2-α-D-GalA-1,4-D-Xyl. It has also been suggested that REO may act as a xylan elongation terminator (Peña et al., 2007). However, the typical polysaccharide direction of synthesis, and the synthesis of xylan observed *in vitro* are both towards NRE, strongly supporting the hypothesis that the REO acts as a primer (Scheller and Ulvskov, 2010; Urbanowicz et al., 2014).

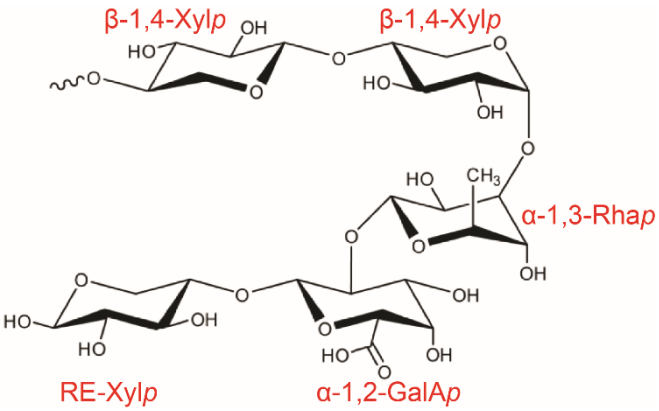


Figure 1.12: Xylan reducing-end oligosaccharide structure.
Adapted from Grantham (2016).

1.3.3 Elongation

The Xyl of the xylan backbone are transferred by GTs from a substrate with high energy (UDP-Xyl,
Figure 1.13). UDP-Xyl is synthesised by the dehydrogenation-oxidation of UDP-Glc into UDP-GlcA, which is then decarboxylated by UDP-D-xylose synthase, irreversibly converting UDP-D-GlcA into UDP-D-xylose (Reiter and Vanzin, 2001). UDP-D-GlcA decarboxylation steps occur in the cytoplasm and in the Golgi apparatus (Rennie and Scheller, 2014). The UDP-Xyl made in the cytoplasm can be transported into the Golgi, but the relative fluxes of the two pathways are unknown.

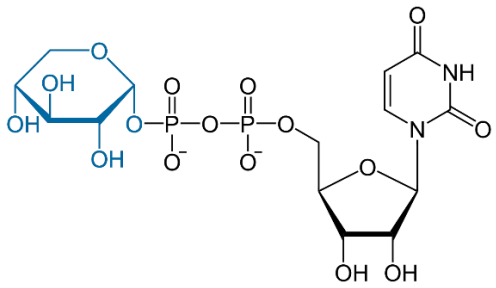


Figure 1.13: Structure of UDP-Xyl
Xyl donor for the synthesis of xylan and other polysaccharides. In blue is highlighted the xylosyl residue.

Three pairs of Golgi-localised GTs are required for the transfer of Xyl from UDP-Xyl to the nascent xylan chain: IRREGULAR XYLEM-9 (IRX9, At2g37090) and IRX9-L (At1g27600); IRX10 (At1g27440), and IRX10-L (At5g61840); and IRX14 (At4g36890) and IRX14-L (At5g67230) (Wu et al., 2010, 2009). In the lack of any paralogue pair (*irx9/9-l*, *irx10/10-l* and *irx14/14-l*), the plant produces traces or no detectable xylan (Brown et al., 2009; Wu et al., 2010, 2009). Knockout mutants for any of these pairs show a severely dwarfed growth phenotype and are sterile. Mutants with a single functional version of any pair member (*irx9*$^{+/-}$*irx9-L*, *irx9irx9-L*$^{+/-}$, *irx10*$^{+/-}$*irx10-L*, *irx10irx10-L*$^{+/-}$, *irx14*$^{+/-}$*irx14-L*, *irx14irx14-L*$^{+/-}$) are capable of producing xylan and grow to WT height and are not sterile (Brown et al., 2009; Keppler and Showalter, 2010; Peña et al., 2007; Wu et al., 2010, 2009). This indicates that to some extent, the members of each of the pairs have a similar function, and originated from gene duplication events.

The degree in which the similar proteins are interchangeable is not fully understood. The xylan backbone is shorter in *irx9*, *irx10* and *irx14*, where is made by a machinery containing only IRX9-L, IRX10-L or IRX14-L respectively (Brown et al., 2009, 2007; Peña et al., 2007). These mutants also possess characteristic collapsed xylem vessels, and *irx9* is slightly dwarfed. In *Arabidopsis*, the inactivation of IRX9 or IRX14 does not cause significant changes of expression of the cell wall synthesis machinery (Faria-Blanc et al., 2018). In contrast, the inactivation of IRX9 or IRX14 homologues in aspen stimulates growth by reprograming the expression of transcription factors associated with cell wall development (Ratke et al., 2018). It was hypothesised that for secondary-cell-wall xylan synthesis, the lack of IRX9/9-L, IRX10/10-L or IRX14/14-L, can be partially compensated by the overexpression of the respective IRX-L genes (Wu et al., 2010, 2009), restoring wild-type (WT) height and producing xylan. Although a restoration of the plant growth and regularity of xylem vessels in the overexpressors was shown, this work did not investigate any recovery of xylan structural properties such as length or arrangement of the substitutions on the backbone. These analyses were carried out in this thesis and are shown in Chapter 3.

The synthesis of xylan in primary and secondary cell walls seems to be done by different sets of enzymes (Figure 1.14). In *Arabidopsis*, IRX9-L and IRX10-L are required for the production of primary-cell-wall xylan, while IRX9 and IRX10 do not seem involved in this synthesis (Mortimer et al., 2015). Transcript analysis in aspen showed a differential expression of IRX9-L homologues in a broad range of primary cell wall producing cells (Ratke et al., 2015). In contrast, IRX9 homologues showed expression mostly in secondary-walled cells. Asparagus IRX9-L homologue showed high expression at the top of the spear, where IRX9 expression is very low (Song et al., 2015). IRX14 and IRX14-L are both required for primary- and secondary-cell-wall xylan, and since they do not have distinct orthologues in other species they are probably the most recent duplication of the three IRX protein pairs in *Arabidopsis* (Jiang et al., 2016; Mortimer et al., 2015; Ratke et al., 2015). Jiang et al. (2016) showed that IRX10 and IRX14 may interact based on heterologous expression in yeast and immunoblot analysis. Additionally, Förster-resonance-energy-transfer (FRET) analysis showed that IRX9-L, IRX10-L and IRX14 may interact (Sorieul et al., unpublished). The evidence suggests that, like the CSCs, the xylan synthesis complex (XSC) differs in cells producing primary and secondary cell walls. The primary cell wall XSC would be composed by IRX9-L, IRX10-L and IRX14, and the secondary cell wall XSC by IRX9, IRX10 and IRX14. Then, in *Arabidopsis*, due to the recent duplication of IRX14, IRX14-L is also involved in both complexes, but its expression can be about a fifth of IRX14 in stems, and a half in other tissues (RNA-Seq data, TRAVA database).

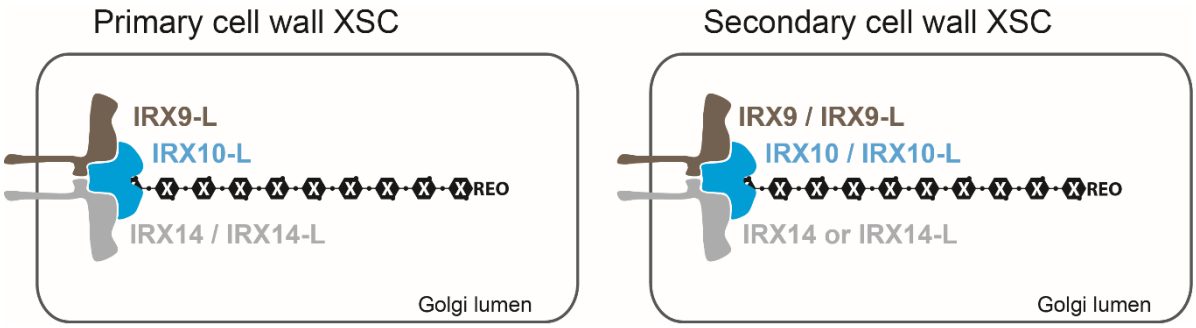


Figure 1.14: Scheme of a simplified versions of putative xylan synthesis complexes from cells synthesising primary and secondary walls.
The lack of IRX9-L or IRX10-L in the xylan synthesis complex (XSC) disrupts the synthesis of xylan in primary cell walls. In secondary-cell-wall xylan production, at least one member of each pair is required for synthesis. X – Xylosyl residue.

IRX9, IRX9-L IRX14 and IRX14-L belong to the GT43 protein family, while IRX10 and IRX10-L are from GT47 family, all inverting GTs (Cantarel et al., 2009). GT43 proteins are type II membrane proteins anchored by a single N-terminal transmembrane domain, and have their catalytic domains facing the Golgi lumen. The GT47s have an additional proteolytic site between the predicted transmembrane and the active site domains. IRX14/14-L and IRX10/10-L have a characteristic DxD motif which is conserved and necessary for the binding of a glycosidic substrate (Ren et al., 2014; Zeng et al., 2016). On the other hand, IRX9/9-L have lost this function, and therefore, possibly they may have a structural role in the XSC. The function of these GT43 proteins in the XSC is not understood. In contrast, IRX10 and IRX10-L have been independently shown to be catalytically active *in vitro*, where they add Xyl to the NRE of xylooligosaccharides (Jensen et al., 2014; Urbanowicz et al., 2014). In the reaction, IRX10 and IRX10-L were able to extend the chain by one Xyl at a time when supplied with UDP-Xyl.

Two additional proteins have been identified to have an effect on xylan synthesis. IRX15 (At3g50220) and its homologue IRX15-L (At5g67210) have been shown to be necessary for a correct cellulose and xylan deposition (Brown et al., 2011; Jensen et al., 2011). In *irx15-l* mutant the xylose levels are unaffected, and *irx15* shows a small but significant decrease in xylose. However, *irx15/15-l* exhibited 65% of xylose levels compared to WT. Xylan extracted from *irx15/15-l* has a higher elution time from size exclusion chromatography, which indicates that this mutant has shorter xylan molecules (Brown et al., 2011). The amount of total xylan in *irx15/15-l* is about half of the WT, but has no difference in the proportion GlcA:Xyl. IRX15 and IRX15-L possess a DUF579 domain, which has no structural similarity to any glycosyltransferases in the Carbohydrate-active enzymes (CAZy) database. It shows homology to xylan methyltransferases (Urbanowicz et al., 2012) but its function in xylan synthesis remains unclear. Since IRX15/15-L are Golgi localised (Jensen et al., 2011), they could be involved in xylan backbone elongation by interacting with the XSC.

1.4 Xylan modifications

After the xylan backbone is synthesised, it can be modified with several substitutions, which vary depending on the plant species, tissue and cell wall type (primary or secondary). In eudicots, glucuronoxylan is mostly decorated with *O*-acetyl (Ac) and O-2-linked (methyl)glucuronic acid substitutions ([Me]GlcA) (Figure 1.15). In monocots, arabinoxylan is decorated mainly with Ara*f*, but Ac and [Me]GlcA are also present. Gymnosperm xylan has O-2-linked MeGlcA and O-3-linked Ara*f*, but no Ac have been reported (Busse-Wicher et al., 2016b). Ara*f* are reported to be O-2-linked to backbone Xyl in eudicots and non-gramineous species (Carpita, 1996; Darvill et al., 1980). *Plantago* (eudicot) xylan backbone is reported to have O-2- and O-3-linked Ara*f* (Guo et al., 2008). In grasses Ara*f* is mostly found O-3-linked, and O-2 linked is reported in seeds (Ebringerová and Heinze, 2000; Peña et al., 2016).

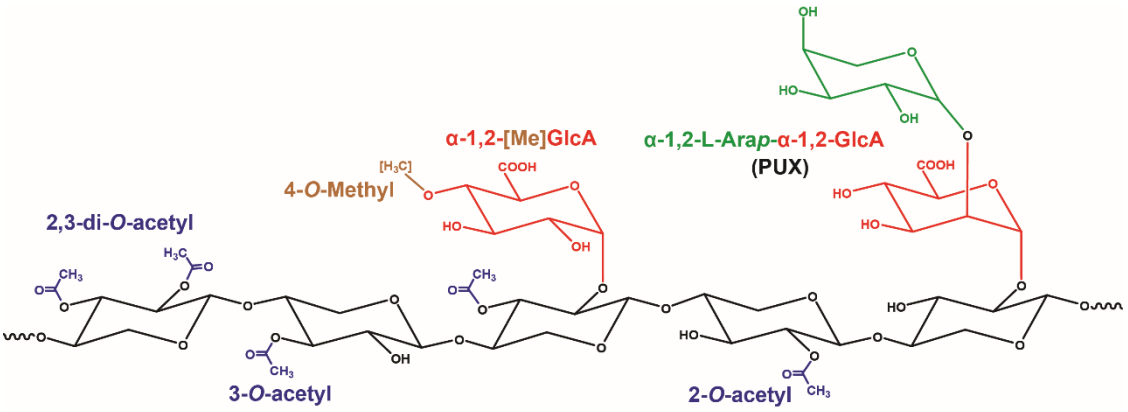


Figure 1.15: Common xylan modifications in eudicots.
Schematic representation of the most common Xyl substitutions. From left to right: 2,3-di-*O*-acetylated Xyl (X^{aa}); 3-*O*-mono-acetylated Xyl (X^{a}); (3-*O*-Ac)(α-[Me]GlcA-1,2)-Xyl ($U^{[me],a}$); 2-*O*-mono-acetylated Xyl (X^{a}); (α-1,2-L-Ara*p*-α-1,2-GlcA)-Xyl (P). Less common 1,2-Ara*f* (A^{2}) and 1,3-Ara*f* substituted Xyl (A^{3}) are annotated A in this thesis for simplicity. Unidentified pentosyl-substituted Xyl are annotated N. Adapted nomenclature from Fry et al. (1993).

1.4.1 Xylan acetylation

Acetate groups are by far the most abundant substitution in eudicot xylan, but the function of this modification in not fully understood. Ac are a more efficient use of carbon than sugars to use in the wall, as these groups are much smaller than monosaccharides (Pauly et al., 2013). Ac can also restrict the accessibility of carbohydrate breakdown enzymes during pathogenic attacks and rumen breakdown of cell walls (Biely et al., 1986). Secondary-cell-wall xylan Ac are necessary and sufficient to keep xylan functional, soluble, and interacting with cellulose (Busse-Wicher et al., 2014; Mortimer et al., 2010).

O-acetyl groups on xylan can be at O-2 and/or O-3 positions on Xyl (Figure 1.15). When a xylan backbone Xyl is mono-substituted with Ac, the substitution can spontaneously migrate in solution between positions O-2, O-3 and O-4 (Figure 1.16) (Kabel et al., 2003; Puchart and Biely, 2015; Urbanowicz et al., 2014). Migration of Ac within a Xyl is only possible when a free hydroxyl is available.

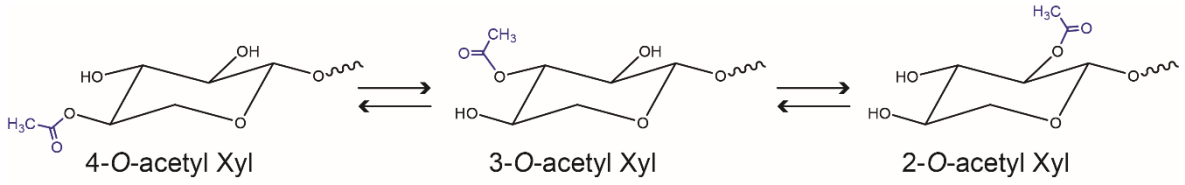


Figure 1.16 Spontaneous migration of Ac within a xylosyl residue.
Migration is possible only when there is an available free hydroxyl on Xyl. Migration to O-4 can occur only in NRE terminal Xyl. Ac migration cannot occur on di-substituted Xyl that are not terminal.

Xylan, like other acetylated polysaccharides, is synthesised in the Golgi, where Ac are transferred from Ac-CoA (Figure 1.17) (Gille and Pauly, 2012). Ac-CoA is thought to be internalised to the Golgi apparatus by a family of four membrane proteins called REDUCED WALL ACETYLATION (RWA1, RWA2, RWA3 and RWA4) (Manabe et al., 2011). This family was discovered by sequence homology to the C-terminus of Cas1p, which is a fungal membrane protein necessary for polysaccharides *O*-acetylation (Anantharaman and Aravind, 2010; Janbon et al., 2001; Manabe et al., 2011). The name RWA is based on the mutant phenotypes, which have a general reduction in the acetylation of all cell wall polysaccharides. When three RWA proteins are inactivated, xylan acetylation can be reduced up to about 50% (Manabe et al., 2013). The quadruple mutant has the most severe phenotype, does not develop stems and shows several undifferentiated tissues.

Xylan acetylation during synthesis is not understood, and prior to this thesis only a single xylan-active acetyltransferase was described. Thus, one of the objectives of this thesis was to investigate additional acetyltransferases that are responsible for the acetylation of xylan.

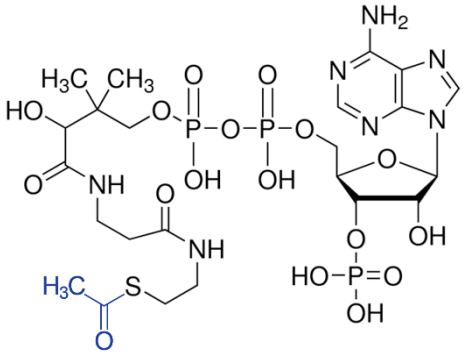


Figure 1.17 Acetyl-CoA structure.
In blue is highlighted the Acetyl group that is transferred to the substrate in acetylations reactions.

1.4.1.1 TBL protein family

A family of putative carbohydrate acetyltransferases was identified in plants by sharing homology to the N-terminus of Cas1p (Anantharaman and Aravind, 2010; Janbon et al., 2001). The N-terminus of Cas1p has sequence similarity to DUF231, which is a plant-specific motif (Bischoff et al., 2010a, 2010b). Firstly, TRICHOME BIREFRINGENCE (TBR) and TRICHOME BIREFRINGENCE-LIKE-3 (TBL3) were identified by having the DUF231 domain and were predicted to have a role in pectin esterification. The *Arabidopsis* mutants *tbr* and *tbl3* were found to have less crystalline cellulose on the secondary cell wall but there was no reduction in the total amount of cellulose (Bischoff et al., 2010a; Potikha and Delmer, 1995). Other homologues were also found to have DUF231 domain in *Arabidopsis* by sequence homology to TBR and TBL3, with a total of 46 members in the TBL family (Bischoff et al., 2010a; Xin et al., 2007). All members of the TBL family have an N-terminal transmembrane domain, a C-terminal DUF231 domain, and a plant-specific TBL domain (Bischoff et al., 2010a). The DUF231 and TBL domains, each possesses a motif that is highly conserved in the GDSL-family of esterases and lipases (Akoh et al., 2004). The TBL domain has a conserved GDSL peptide, and DUF231 has a highly conserved amino acid sequence (DCXHWCLPGXXDXWN) towards the C-terminus of the proteins with a DxxH motif. The Ser residue in the GDSL motif, with Asp and His residues in the DxxH motif were determined to form the catalytic triad in the crystal structure of a fungal rhamnogalacturonan acetylesterase (Mølgaard and Larsen, 2004).

Some TBL proteins have been identified to be involved in acetylation of different polysaccharides of the wall. AXY4/TBL27 and AXY4-L/TBL22 have been shown to be required for acetylation of xyloglucan (Gille et al., 2011). Recently, TBL23, TBL24, TBL25 and TBL26 have been recently shown to be capable of transferring Ac to mannohexaose *in vitro* (Zhong et al., 2018c), and TBL10 was shown to be required for acetylation of pectin (Stranne et al., 2018). As mentioned before, xylan is a highly acetylated polysaccharide and its Ac are displayed in several configurations (Figure 1.15), however, only one acetyltransferase had been previously shown to be active on this polymer at the start of the work for this thesis; ESK1 (ESKIMO1/XOAT1/TBL29) (Urbanowicz et al., 2014). Other possible xylan acetyltransferases from the TBL family were explored in depth in Chapter 6 of this thesis.

1.4.1.2 ESK1

ESKIMO1 (ESK1, At3g55990) was initially identified by screening for freeze tolerance in a population of *Arabidopsis* wild-type mutagenised with ethyl methanesulfonate, (Xin and Browse, 1998). *esk1* mutant was found to be resistant to freezing temperatures, which gave origin to the name. It was later on that ESK1 was found to have the TBL and DUF231 domains, associating it with the TBL protein family (Bischoff et al., 2010a).

esk1 mutants have, collapsed xylem vessels, a mild growth phenotype with a dark green colour, and low water flux which is characteristic of xylan-deficient mutants (see Figure 1.18) (Bouchabke-Coussa et al., 2008; Brown et al., 2009, 2007; Lefebvre et al., 2011; Peña et al., 2007; Xin and Browse, 1998). The reduced water transport is thought to be the cause for abscisic acid accumulation. *esk1* shows a typical metabolome of plants which have been exposed to drought, salt or osmotic stress, having constitutively activated genes for response to oxidative stress (Lugan et al., 2009). The increased stress tolerance shown by *esk1* is likely a consequence of the irregular xylem phenotype and low water transport displayed by these plants (Bensussan et al., 2015; Bouchabke-Coussa et al., 2008; Lefebvre et al., 2011; Lugan et al., 2009). However, a recent report shows that *esk1* plants have constitutively activated the biosynthetic pathway for the stress-related phytohormone strigolactone, and when the pathway was blocked, the tolerance to cold is lost (Ramírez et al., 2018). In contrast to what has been previously proposed, the authors suggest that in WT, cell wall defects are somehow ‘sensed’ by the plant, triggering a stress response, and that this response is constitutively activated in *esk1*. However, they did not show evidence for cell wall sensing mechanisms, nor stress resistance caused directly by defects on the wall.

Figure 1.18 *esk1* mutant growth and collapsed xylem vessels phenotypes.
Panel A was reproduced from Lefebvre et al. (2011) and shows growth of *Arabidopsis* WT (left) and *esk1* (right) grown for 30 days at 22 Celsius degrees (°C). Panel B was reproduced from Yuan et al. (2013) and shows cross-sections of xylem bundles of stems. Arrows show a mild deformation of xylem (xy) vessels in *esk1*.

ESK1 is a 57 kDa protein constituted by 487 amino acids that is expressed in vasculature of stems, in cells under cell wall thickening (Lugan et al., 2009; Xin et al., 2007). As mentioned before, ESK1 is the member number 29 of the TBL family (TBL29), and possesses all of the characteristic domains of the TBL family. It has an N-terminal transmembrane domain, followed by the TBL domain that extends from residues 139 to 227, and the DUF231 domain on the C-terminus from residues 310 to 483 (Xin et al., 2007; Zhong et al., 2018b).

ESK1 has been characterised as a xylan-specific acetyltransferase. It was initially thought that ESK1 was involved in polysaccharide acetylation, due to its similarity to Cas1p, and the reduction in ester linkages observed in *esk1* cell wall (Gille and Pauly, 2012; Lefebvre et al., 2011). Then, 2D-nuclear magnetic resonance (NMR) experiments on *esk1* alcohol-insoluble residue (AIR) suggested that the reduction in acetate groups in *esk1* is specific for xylan and/or mannan mono-acetylations (Xiong et al., 2013). Later, *in vitro* activity experiments showed that ESK1 can acetylate xylooligosaccharides (but not mannooligosaccharides) with up to three Ac confirming ESK1 specificity for xylan (Urbanowicz et al., 2014). The *in vitro* activity also revealed that ESK1 acetylates Xyl only at O-2, since 3-*O*-acetyl groups were not detected in the products before 30 min of incubation. The delayed detection of 3-*O*-Ac is consistent with the spontaneous migration of Ac within backbone residues (Kabel et al., 2003; Pauly, 1999). Consistent with ESK1 regiospecificity (position of the transferred Ac) at

O-2, Grantham (2016) showed 2-D NMR spectra of *esk1* xylan which suggested that the mutant xylan lacks mostly O-2-linked Ac on the backbone Xyl. His data also suggested that esk1 has a reduction of 2,3-di-acetylated Xyl.

Consistent with the reported ESK1 specificity for xylan (Xiong et al., 2013) several later publications from a different research group showed a similar reduction in *esk1* overall xylan acetylation (Yuan et al., 2016c, 2016a, 2016b, 2013). Consistently, their 1-D NMR spectra of *esk1* xylan showed reduced O-2 and O-3-linked mono-acetylated Xyl. However, they attributed the lack of 3-*O*-Ac to the 3-*O*-acetyltransferaseactivity of ESK1, arguing that spontaneous migration does not occur in their experimental conditions. However, the experimental conditions that they used are based on a previous report (Evtuguin et al., 2003) which stated that Ac migration may occur in such conditions. The result is also in contrast to the *in vitro* findings of Urbanowicz et al. (2014).

1.4.1.3 AXY9

An additional player in xylan acetylation pathway has been recently discovered, and its activity was proposed to act between RWA and ESK1. ALTERED XYLOGLUCAN9 (AXY9, At3g03210) was found by screening for xyloglucan acetylation deficiencies in an methanesulfonate-mutagenised *Arabidopsis* population (Schultink et al., 2015). *axy9* mutants exhibit less acetyl-ester linkages; up to 80% reduction in xylan acetylation and 60% in xyloglucan acetylation with respect to wild-type levels. The mutants' growth is severely dwarfed, they have a dark green colouration, and collapsed xylem vessels. AXY9 phenotype is similar to the exhibited by other mutants with reduced acetylation in the wall (i. e. *rwa1/2/3/4*; Manabe et al., 2013). AXY9 protein is distinct from the TBLs, but does share homology with them and with other acetyltransferases, suggesting that AXY9 may act to produce an unknown acetylated intermediate between RWA and the TBLs, being part of a general polysaccharide acetylation pathway in the Golgi (Schultink et al., 2015).

1.4.2 Xylan glucuronidation

After acetate, glucuronic acid is the most abundant substitution on xylan in eudicots. In secondary cell walls, on average around one in every eight backbone Xyl carries a [Me]GlcA (Brown et al., 2007; Rennie and Scheller, 2014). Xylan [Me]GlcA are exclusively α-1,2-linked to the backbone Xyl (Aspinall, 1980), and the spacing between two [Me]GlcA substitutions varies along the polymer (Bromley et al., 2013). The glucuronidated Xyl of secondary-cell-wall xylan are normally also 3-*O*-acetylated (Evtuguin et al., 2003; Teleman et al., 2002), but small amounts of [Me]GlcA-Xyl that are non-acetylated have also been described (Chong et al., 2014). In woody plants, all xylan GlcA substitutions are *O*-methylated at C-4, but in *Arabidopsis* only about 60% of the xylan GlcA are *O*-methylated (Li et al., 2013; Peña et al., 2007). The proportion of MeGlcA:GlcA varies in different plant growth conditions and in different portions of the stem (Faria-Blanc, 2014).

1.4.3 GlcA methylation

The function of the *O*-4-methyl groups on GlcA is not fully understood. Because [Me]GlcA substitutions may link to lignin (Figure 1.19) (Imamura et al., 1994; Watanabe and Koshijima, 1988), possibly the methyl group facilitates this link by contributing to the hydrophobicity of the environment around the carboxylic acid. Possibly a H-bond between the carboxylic acid and the hydroxyl at O-4 may be hindered by the methylation of the latter. In mutants that lack xylan [Me]GlcA substitutions (*gux1/2*, see section 1.4.3.1) the xylan is more efficiently extractable (Lyczakowski et al., 2017), supporting the hypothesis of the link between xylan and lignin.

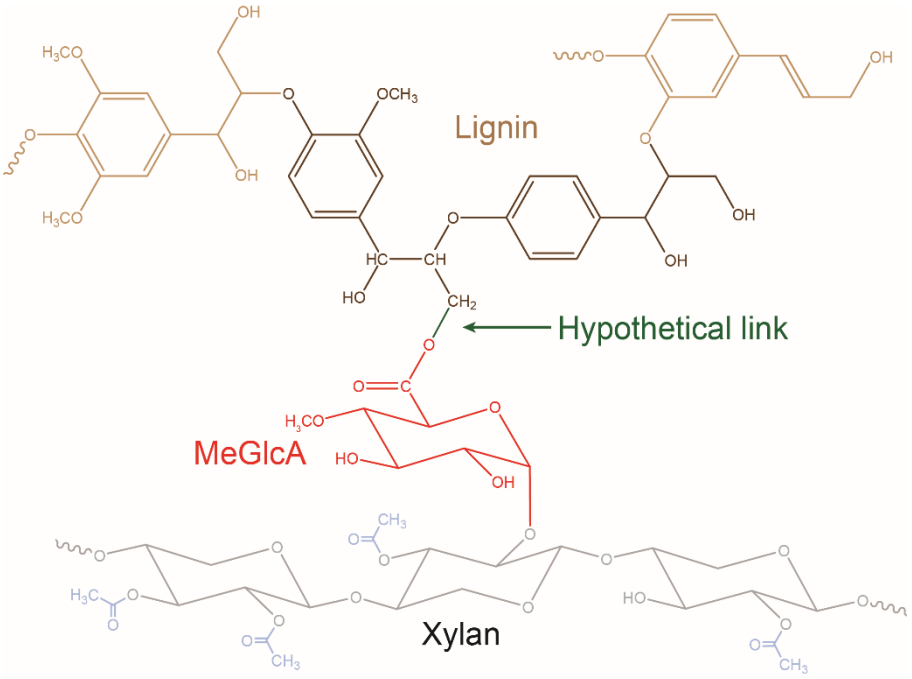


Figure 1.19: Proposed link between xylan and lignin.
An ester link (green) between GlcA of xylan (red) and a lignin residue (brown) is shown. A coniferyl alcohol is illustrated to exemplify the bond, but *p*-coumaryl and sinapyl residues could also form this link. The hypothetical bond shown is on the γ carbon of the monolignol, but the α carbon may also form this bond. (Terrett and Dupree, 2019; Watanabe and Koshijima, 1988).

The methylation on the GlcA is transferred by three glucuronoxylan methyltransferases (GXM1, GXM2 and GXM3) using S-adenosyl-L-methionine (SAM) (Kauss and Hassid, 1967; Li et al., 2013; Urbanowicz et al., 2012; Yuan et al., 2014). GXM1 alone is responsible for 75% of xylan GlcA methylation, and *gxm1/2/3* mutant has nearly undetectable MeGlcA (Urbanowicz et al., 2012; Yuan et al., 2014).

In *Arabidopsis*, the proportion of xylan GlcA that is methylated *in vivo* seems to be affected by other processes in the synthesis pathway. In *irx9* and *irx14* mutants, which have less and shorter xylan, the 1:8 ratio of GlcA:Xyl is maintained, but a higher proportion of GlcA are methylated (Peña et al., 2007). A saturation mechanism was proposed in which methylation may keep pace with xylan elongation. If the speed in which xylan is synthesised decreases, the rate at which GlcA side chains become available substrates for methylation consequently decreases. Acetylation deficient mutants, *rwa* and *esk1* exhibit a significant increase in the total amount of [Me]GlcA on xylan respect to WT (Grantham, 2016; Grantham et al., 2017). However, the proportion MeGlcA:GlcA in *rwa* mutants is decreased while in *esk1* this proportion is increased (Grantham, 2016; Grantham et al., 2017; Manabe et al., 2013; Yuan et al., 2014). Therefore, the saturation of GXM activity may affect how many GlcA become methylated, but the acetylation context around GlcA substitutions in *esk1* and *rwa* also affects the activity of GXM but in opposite manners. This implies that GXM recognise Ac which do not correspond with ESK1 regiospecificity (O-2). Also, this evidence strongly suggests that multiple xylan acetyltransferase activities are possible, which impact differently on GXM activity. The *in vitro* activity of GXM1 has been shown to be several fold faster when it acts on acetylated xylan than on non-acetylated xylan (Faria-Blanc, 2014). This implies that the methylation of the GlcA is downstream from the backbone acetylation stage. It has not been previously shown *in vivo* which Ac position/context on the backbone impacts on the methylation of GlcA. The effect of the 3-*O*-acetylation of GlcA-bearing Xyl on the downstream process of methylation of the GlcA was investigated *in vivo* in this thesis (Chapter 6).

1.4.3.1 GUX enzymes

GlcA is transferred to the xylan backbone by retaining glucuronyltransferases from the GT8 family (Rennie et al., 2012). While GLUCURONIC ACID OF XYLAN 1 (GUX1, At3g18660) and GUX2 (At4g33330) glucuronidate secondary-cell-wall xylan, GUX3 (At1g77130) has been shown to be responsible only for primary-cell-wall-xylan [Me]GlcA substitutions (Figure 1.20) (Mortimer et al., 2015, 2010; Rennie et al., 2012). The co-inactivation of GUX1 and GUX2 in *Arabidopsis* (*gux1/2* mutant) produces almost a complete depletion of GlcA on xylan from mature portions of stems, while single knockouts *gux1* and *gux2* display partial depletions of GlcA (Bromley et al., 2013; Mortimer et al., 2010). *gux1* mutant xylan has a GlcA frequency reduction to about 30% of WT, while *gux2* has a significant reduction to about 80% of WT (Mortimer et al., 2010). The absence of GUX1 and GUX2 are therefore additive. The overexpression of GUX1 or GUX2 in the absence of the other does not significantly increase the amount of [Me]GlcA substitutions on xylan, indicating that their activities are limited to a specific region of xylan (Bromley et al., 2013). The total acetylation on xylan does not seem to be substantially affected by the inactivation of GUX1 and GUX2 (Busse-Wicher et al., 2014).

In *gux1/2* where only GUX3 is active, tissues enriched with primary cell walls show a correlated enrichment of GlcA substitutions on xylan (Mortimer et al., 2015), attributing GUX3 activity to primary-cell-wall xylan. GUX3-dependent GlcA decorations can also be methoxylated at C-4. Additionally, in newly-grown stem xylan, a 1,2-linked pentopyranosyl residue was found on some non-methylated GlcA substitutions (Mortimer et al., 2015). The authors proposed that this pentosyl residue (Pent) may be an α-L-arabinopyranose (Ara*p*), but they could not confirm the identity, and named PUX the pentosylated GlcA sidechain. Later, Peña et al. (2016) confirmed the identity of Ara*p* by 2D-NMR on GlcA in a similar structure in xylan from monocot glucuronoarabinoxylan. Xylan from callus was shown to have only GUX3-dependent [Me]GlcA, and possibly PUX sidechains (Mortimer et al., 2015). The protein responsible for the transfer of Ara*p* to the GlcA has not yet been reported.

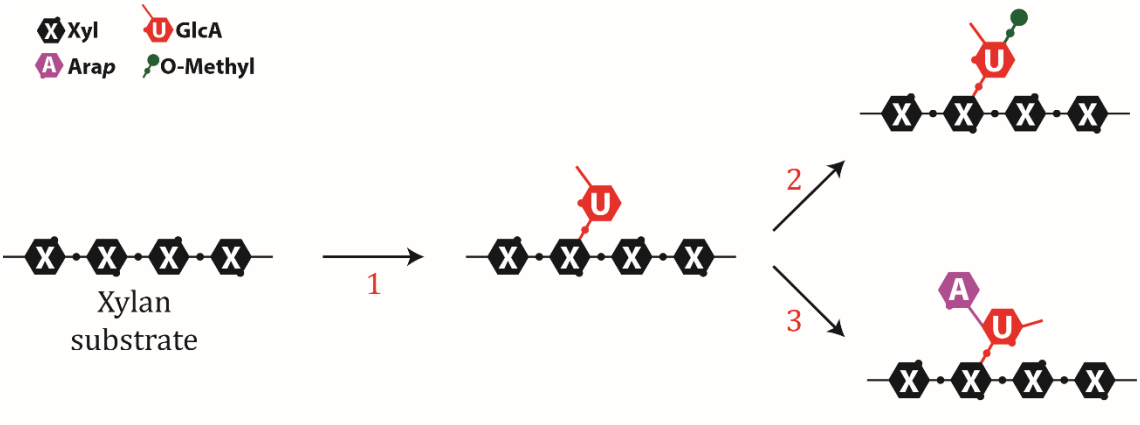


Figure 1.20: GlcA by *Arabidopsis* GUX activities.
GUX1, GUX2 and GUX3 are responsible for the transfer of α-1,2-GlcA onto the xylan backbone Xyl (1), which can be further methylated at C-4 (2). Only the GlcA transferred by GUX3 to primary-cell-wall xylan can be further modified with α-1,2-Ara*p* (3).

1.4.4 Pattern of substitutions on the xylan backbone

The function of xylan relates to the molecular interactions that it forms with cellulose and other components of the wall. These interactions are possible because of the specific type and distribution of the substitutions on the backbone (Grantham et al., 2017). The patterns consist of specific spacing between xylan modifications. In all analysed land plants, the major substitutions that xylan possesses form particular patterns along the chain. In eudicots, Ac and [Me]GlcA form this pattern (Bromley et al., 2013; Chong et al., 2014; Mortimer et al., 2015). Gymnosperms and lower land plants also exhibit specific patterns of their xylan substitutions (Busse-Wicher et al., 2016b).

Since the function of xylan is dependent on the pattern of its decorations, the description of the pattern is informative towards understanding the role of xylan in the wall. In *Arabidopsis*, the pattern of [Me]GlcA substitutions is known (section 1.4.4.2), but the arrangement of Ac in different portions of the xylan is poorly understood. One of the focuses of this thesis is the description of different acetylation patterns on xylan (Chapter 4 and Chapter 5).

1.4.4.1 Acetylation pattern of xylan

The main xylan acetylation sequence on the backbone (Figure 1.21) was determined in eudicot xylan by analysing the product of a partial xylan digestion with endoxylanases (Busse-Wicher et al., 2014; Chong et al., 2014). The predominant products from both WT and *gux1/2* were xylooligosaccharides with alternating Ac, indicating that mostly every-other Xyl has one acetate group. The inactivation of GUX1 and GUX2 showed no alteration of the main alternating pattern of acetylation, indicating that the process of xylan glucuronidation does not affect the xylan acetylation patterning.

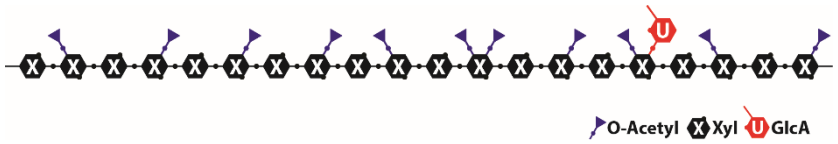


Figure 1.21: Main alternating acetylation pattern of xylan.
Every other Xyl from the xylan backbone is acetylated. Acetylated Xyl are mostly mono-acetylated, but some are di-acetylated or co-substituted with [Me]GlcA. Any arrangement of de-acetylated residues with respect to [Me]GlcA or position in the molecule is unknown.

ESK1 was shown to be responsible for the alternating acetylation pattern, and *esk1* shows a disruption of it (Grantham, 2016). The Ac in *esk1* are mostly nearby the [Me]GlcA substitutions or isolated in long undecorated extensions. In contrast, *rwa* mutants, which also have a reduction in xylan acetylation, have gaps between the single acetates, but the remaining Ac are still ‘in frame’, having an even number of Xyl of spacing between them (Grantham et al., 2017). It is possible that even though in *rwa* there is less Ac-CoA in the Golgi, ESK1 has enough substrate to generate the acetylation pattern. Possibly, *rwa1/2/3/4* quadruple mutant lacks enough Ac-CoA in the Golgi for ESK1, being incapable of producing a functional xylan, as well as other essential acetylated components of the wall.

The number and distribution of Ac on xylan vary among different species (Busse-Wicher et al., 2016b; Carpita, 1996; Scheller and Ulvskov, 2010). The high levels of acetylation on xylan and the characteristic alternating pattern may act to prevent aggregation of xylan chains or xylan-coated cellulose microfibrils (Busse-Wicher et al., 2014; Pawar et al., 2013). In *Arabidopsis* secondary-cell-wall xylan, the number of Ac per Xyl (DA, degree of acetylation) is around 0.60 (Busse-Wicher et al., 2014; Grantham, 2016; Xiong et al., 2013). Considering the main alternating pattern of acetylation, a DA>0.5 suggests that other Ac patterns are possible which may include 2,3-di-Ac-Xyl and/or consecutive mono-Ac-Xyl.

In contrast to secondary-cell-wall xylan, primary-cell-wall xylan acetylation has been detected but the patterning has not been investigated. One publication has described a single xylooligosaccharide structure containing both PUX and Ac linked to the same Xyl ($P^{a}XX$) (Chong et al., 2015). In this publication, the authors postulated that this structure originates from primary cell walls, given its enrichment in *irx9*, however, they could not discard the possibility that $P^{a}XX$ may originate from secondary cell wall. Aside from this report, no acetylation has been reported in primary-cell-wall xylan of eudicot vegetative tissues. If this xylan is acetylated, it raises several questions: What is the function of primary cell wall Ac? What are the positions of the acetates on the Xyl? Are there 2,3-di-*O*-acetylated Xyl? Is there an acetylation pattern? If there is, is it similar to the pattern in secondary-cell-wall xylan, or is it different? Is ESK1 or another TBL responsible for this acetylation? Answering these questions will allow us to understand the function of xylan, and potentially to manipulate xylan structure. Just as it did once xyloglucan synthesis was well characterised. Some of these questions are answered and discussed in Chapter 5 and Chapter 6 of this thesis, and others remain to be clarified.

1.4.4.2 Glucuronic acid pattern of xylan

The way in which GUX1 and GUX2 decorate the xylan not only differs in the amount, but also in the spacing between GlcA substitutions (Bromley et al., 2013). [Me]GlcA transferred by GUX1 are widely distanced one from another, while GUX2 [Me]GlcA are more clustered together. More importantly, almost all GUX1-glucuronidated Xyl are distanced by even number of backbone Xyl (even pattern), with a predominance

of 6, 8, 10 and 12 Xyl of distance between them (Figure 1.22). In contrast, GUX2-GlcA substitutions do not have a preference for even or odd number of backbone Xyl of distance (non-even pattern), and they are most frequently spaced by five, six or seven residues (Bromley et al., 2013). GUX1 and GUX2 appear to be active on distinct regions or domains. The [Me]GlcA pattern of secondary-cell-wall xylan displayed by *gux1* and *gux2* mutants, is only the one that is produced by the respective active enzyme. Moreover, GUX1 and GUX2 do not seem to compensate for the absence of the pattern of the other. The overexpression of GUX2 in WT and *gux1* does not increase the glucuronidation levels on xylan. On the contrary, the overexpression of GUX1 in WT and *gux2* slightly increases the yield, leading to less spacing between [Me]GlcA substitutions, but these newly added GlcA also follow an even pattern (Bromley et al., 2013). The extra GlcA added by GUX1 therefore seems to be limited to the same portion of xylan.

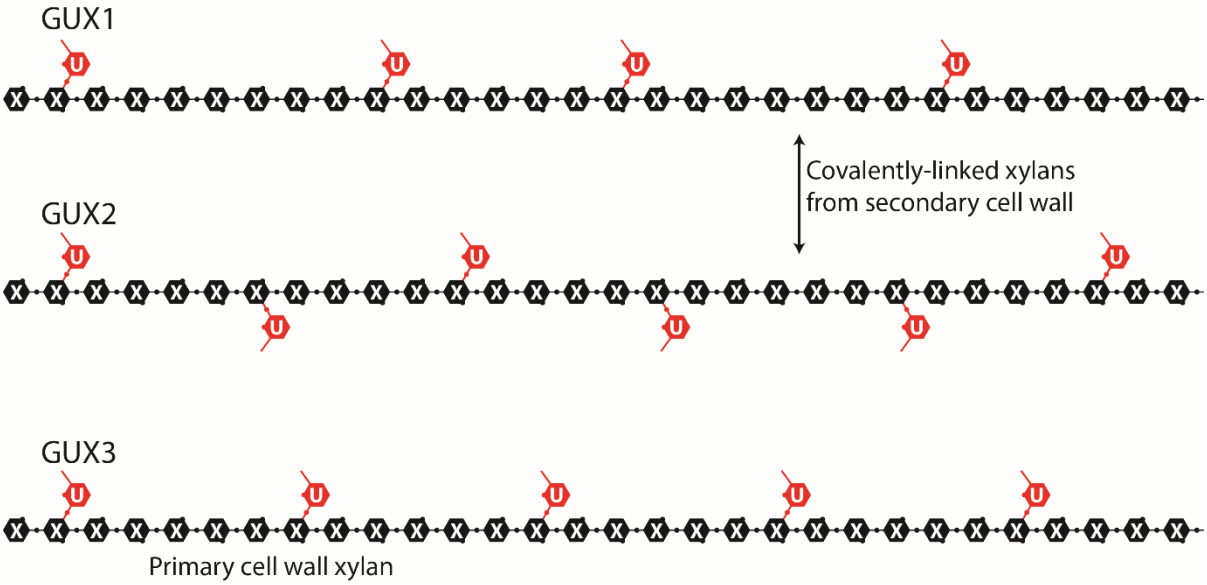


Figure 1.22: GlcA pattern of xylan by GUX enzymes.
Xylan portions are represented in a two-fold screw configuration for simplicity.
The spacing between GlcA (or [Me]GlcA) is characteristic for each GUX activity. GUX1 and GUX3 transfer GlcA substitutions at even number of backbone Xyl of distance from each other. GlcA transferred by GUX3 are constantly every six Xyl, while the spacing varies between GlcA when transferred by GUX1. GUX2 decorates the xylan without a preference for even or odd number of Xyl of distance between GlcA. GUX1 and GUX2 decorate the same xylan molecules for secondary cell wall synthesis, while GUX3 decorates xylan from primary cell walls. X – xylosyl residue, U – glucuronic acid residue.

The inactivation of GUX1 or GUX2 is thought to leave a non-glucuronidated but acetylated region on xylan (Bromley et al., 2013; Busse-Wicher et al., 2014). The portions of xylan that are decorated by GUX1 and by GUX2 are not separable by charge or by size, implying they are covalently linked and contained within the same xylan molecules (Bromley et al., 2013). Nevertheless, the number of GUX1 or GUX2-decorated regions in each xylan molecule, and their distribution along the xylan molecule remains to be determined.

Primary-cell-wall xylan have been shown to have a unique [Me]GlcA pattern. These GUX3-dependent [Me]GlcA decorations follow an even-pattern, similar to the one generated by the activity of GUX1 (Mortimer et al., 2015). However, these [Me]GlcA have been shown to be uniquely distanced by six backbone residues (Figure 1.22). Also, the xylan chains decorated by GUX3 have a smaller size to the ones from secondary cell walls (M Busse-Wicher, unpublished data). Phylogenetic analyses of *Arabidopsis* GUXs and homologous GT8 protein sequences from other higher plant species showed that GUX1 and GUX3 group together in a single clade, while GUX2 is in a separate branch (Busse-Wicher et al., 2016a; Rennie et al., 2012). Perhaps, the amino acid sequence and structure of the GUX enzymes may be related to their activity, as both GUX1 and GUX3 produce a [Me]GlcA even-pattern, while the pattern by GUX2 is not. Alternatively, it may be the acetylation context which determines the pattern produced and the regions of xylan that each GUX acts on.

The acetylation of secondary-wall xylan carried out by ESK1 impacts on the [Me]GlcA patterning (Grantham et al., 2017). The reduced acetylation in *esk1* seems to dysregulate the activity of GUX1, since the even pattern of [Me]GlcA is lost. This effect is not observed in *rwa* mutants, indicating that ESK1 is specifically important for the generation of the even pattern of [Me]GlcA. The dysregulation of GUX2 in *esk1* is not as obvious, because the non-even [Me]GlcA pattern produced by GUX2 is similar to the one observed in *esk1* (Grantham et al., 2017). However, the severe dwarfism exhibited by both mutants *esk1 gux1* and *esk1 gux2*, indicates that their xylan is not functional, and suggests that both GUX1 and GUX2 may be affected by the lack of ESK1. The impact of ESK1 acetylation in the [Me]GlcA pattern, but no change in the acetylation pattern in *gux1 gux2*, indicates that the activity of GUX1 and GUX2 occurs downstream to ESK1 acetylation. Whether ESK1 activity affects primary-cell-wall xylan acetylation has not been shown. If it does, would the lack of ESK1 affect GUX3

activity on xylan, and/or the addition of PUX to GlcA? This is investigated in Chapter 6 of this thesis.

1.4.4.3 Interaction between xylan and cellulose

The patterns of Ac and [Me]GlcA on xylan play a crucial role in the interactions between xylan and cellulose. The molecular structure of cellulose and xylan are relatively similar, being both homopolymers with a repeating backbone of β-1,4-linked pyranosyl residues (Figure 1.5, page 7). All hydroxyl groups of the Xyl are equatorial, which enables the xylan backbone to form intermolecular H-bonds with cellulose chains, analogously to how the glucan chains do so within a microfibril (Busse-Wicher et al., 2016a). For xylan, cellulose and other β-1,4 glycans, the glycosidic dihedral torsion angles Φ and Ψ describe the relative orientation of contiguous backbone residues in the same chain (Figure 1.23), and are defined respectively by (O5-C1-O1-C4') and (C1-O1-C4'-C5'). In crystalline regions of cellulose microfibrils, the sum of Φ and Ψ of the glucan chains is ~120º, which means that the structure of the chain follows a two-fold helical screw (Figure 1.24) (French and Johnson, 2009; Mazeau et al., 2005). Molecular dynamics simulations of naked and acetylated xylooligosaccharides adsorbed onto cellulose microfibrils (Figure 1.25) have shown that xylan may be capable of adopting an analogue two-fold screw conformation, emulating H-bonds with cellulose as if it was a surface glucan chain (Busse-Wicher et al., 2014). Undecorated xylan would therefore be capable of interacting with cellulose hydrophilic surfaces in a two-fold screw configuration, however, the xylan backbone is highly substituted. If xylan decorations were situated in between the two polysaccharides, steric hindrance would impede the interaction between xylan and the hydrophilic cellulose surfaces. For this reason, it is thought that the even pattern of [Me]GlcA, and the alternating acetylation pattern of xylan are present in a region that interacts with cellulose (Grantham et al., 2017).

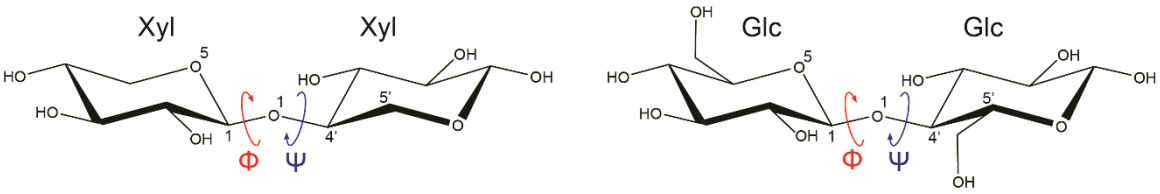


Figure 1.23: Glycosidic bond torsion angles.
Phi (Φ) and Psi (Ψ) angles describe the relative orientation of consecutive sugar residues in a β-1,4-linked pyranosyl chain. If the sum of Φ and Ψ equals 120º, the consecutive residues are facing in opposite directions. A 60º rotation between consecutive residues is indicated by Φ + Ψ being equal to ~50° (right handed) or ~190° (left handed). Diagram based on two publications (French and Johnson, 2009; Mazeau and Heux, 2003).

Figure 1.24 Two-fold and three-fold helical screw configurations of glucan and xylan chains.
Xylan and cellulose glucan chains are found in a two-fold helical screw configuration when forming a crystalline microfibril. Xylan can also adopt a three-fold screw conformation in solution. In the two-fold screw configuration the polymer is flat and linear, and every second residue has the same orientation of the first. In a three-fold screw, every sixth residue has the same orientation as the first. Modified from Dupree et al. (2015).

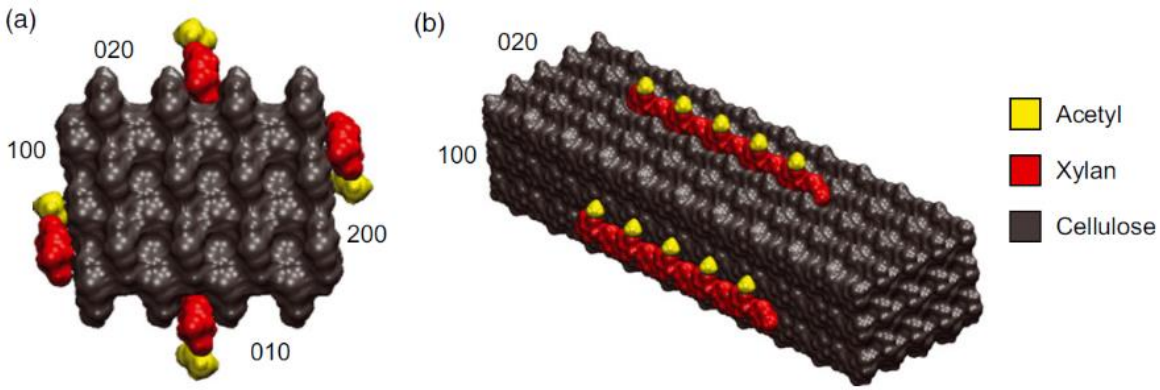


Figure 1.25: 3D model of xylan adsorbed onto a cellulose microfibril.
The end view (a) and side view (b) are shown. DP10 (degree of polymerisation 10) xylan chains with evenly spaced 2-*O*-Ac decorations in alternating Xyl are displayed as a two-fold helical screw. Xylan chains are positioned on hydrophilic (020 and 010) and hydrophobic (100 and 200) faces of cellulose. Modified from Busse-Wicher et al. (2014).

As mentioned above (section 1.4.4.2) GUX1 and GUX2 are thought to glucuronidate the xylan in two discrete zones or domains producing different [Me]GlcA patterns (Bromley et al., 2013). These domains were named compatible and incompatible, according to their hypothetical ability to bind to hydrophilic surfaces of cellulose microfibrils (Figure 1.26). The domain decorated by GUX1 was named compatible, since GUX1 generates the even pattern of [Me]GlcA on xylan which would allow xylan to adopt a two-fold screw configuration, with all [Me]GlcA substitutions facing away from the cellulose microfibril (Busse-Wicher et al., 2014). The incompatible domain is thought to be decorated by GUX2 with a non-even pattern, which causes [Me]GlcA to face facing towards two directions when adopting a two-fold screw configuration, which makes xylan incompatible for this type of interaction. However, incompatible xylan could interact in a two-fold-screw manner to hydrophobic surfaces of cellulose.

A threefold helical screw occurs when the sum of Φ and Ψ equals ~50° or ~190° (French and Johnson, 2009; Mazeau and Heux, 2003). Amorphous cellulose, and xylan in solution, have a three-fold screw structure given their predicted dihedral angles. Both xylan conformations exist in the wall (Dupree et al., 2015). Solid-state NMR experiments showed a similar rigidity between cellulose and two-fold screw xylan, while three-fold screw xylan appeared to be more mobile (Simmons et al., 2016). The NMR experimental approach confirmed the molecular dynamics previously obtained (Busse-Wicher et al., 2014), showing that two-fold screw xylan is in proximity to cellulose, and that cellulose is necessary for the xylan two-fold screw configuration (Simmons et al., 2016).

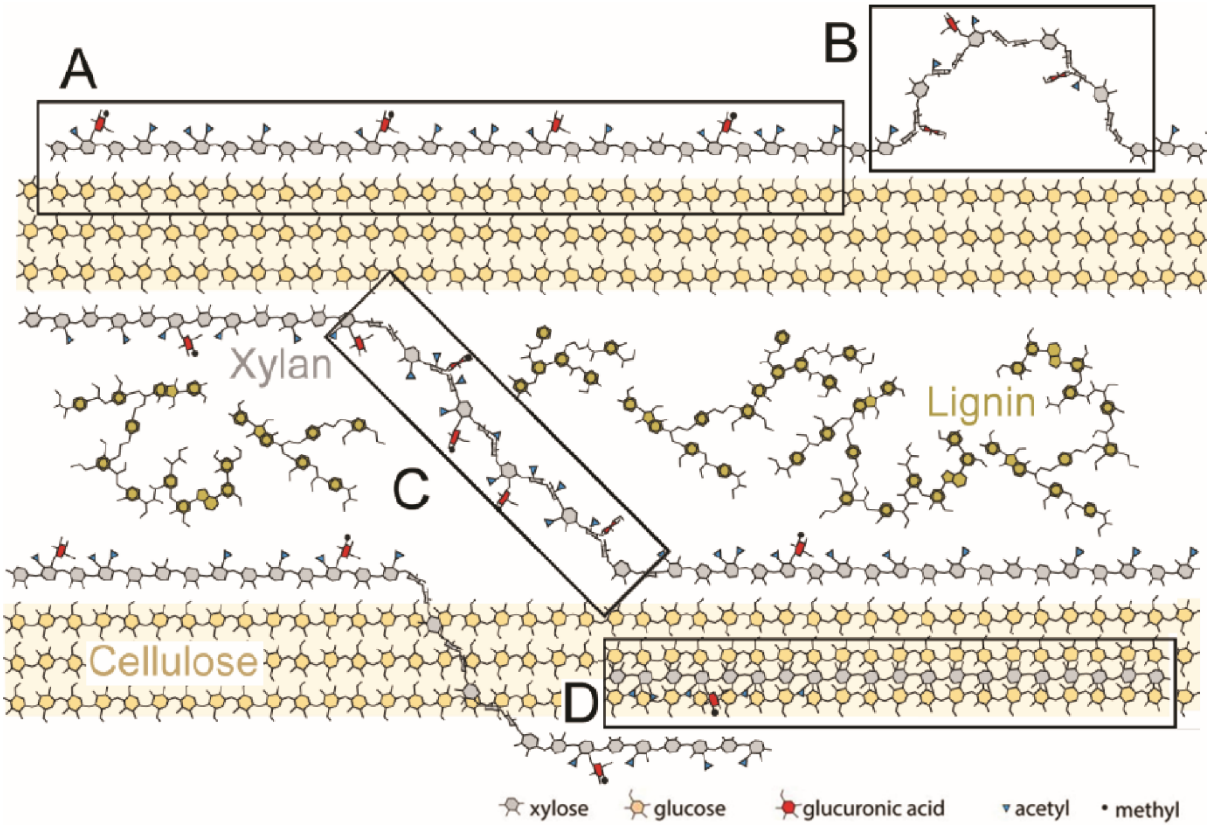


Figure 1.26: Model scheme of possible functions of xylan compatible and incompatible domains in eudicot secondary cell walls.
Compatible xylan could interact with either hydrophilic (A) or hydrophobic (D) surfaces of cellulosed microfibrils, in a two-fold screw configuration. Incompatible xylan can cross-link two microfibrils in a three-fold screw (C) or span one microfibril groove and dock onto another. Incompatible xylan could also be tethered in a three-fold screw near hydrophilic surfaces of microfibrils (B). Modified from Busse-Wicher et al. (2016a).

A xylan-cellulose interaction model was then proposed including the acetylation and the [Me]GlcA patterns on xylan together (Figure 1.26). The model shows the possible cellulose cross-linking function of xylan, which would be driven by the incompatible domain, while the compatible domain would be binding to the hydrophilic surfaces of the microfibrils. The acetylation patterns shown are only hypothetical, as only the main alternating pattern is known (Busse-Wicher et al., 2014; Chong et al., 2014).

The acetylation pattern of xylan is necessary for the interaction between xylan and cellulose microfibrils. If every other xylose is mono- or di-acetylated it allows the xylan chain to adopt a two-fold screw configuration (Figure 1.27). The unpatterned xylan of *esk1* is unable to bind to the hydrophilic surfaces of cellulose (Grantham et al., 2017). Consistent with this hypothesis, no two-fold screw xylan is detected in *esk1* basal stem cell walls by solid-state NMR. In *esk1*, the deficit in xylan acetylation, and the disruption of the acetylation and GlcA patterns possibly makes the xylan partially non-functional, reducing the overall strength of the cell wall by a lack of cross-link connections between cellulose microfibrils. This could explain the collapse of the xylem vessels observed in *esk1* and in other xylan mutants.

Figure 1.27: Model of the effects of the alteration of acetylation pattern and the impact on the xylan binding to cellulose.
In wild-type (left hand side), the XSC first synthesises the xylan backbone in the Golgi apparatus and the activity of ESK1 directs the acetylation of every other Xyl. Next, GUX1 and GUX2 glucuronidate the acetylated xylan, led by the acetylation pattern of xylan. GUX1 adds GlcA distanced by even number of residues, possibly by acting in "gaps" on the acetylation pattern. The xylan is then exported into the matrix and adsorbed onto the cellulose microfibrils. In *esk1* (right hand side), the acetylation pattern is missing, which causes a dysregulation of the activities of GUX1 and possibly GUX2, affecting the compatibility of xylan with the later interaction with cellulose hydrophilic surfaces. Reproduced from Grantham et al. (2017).

1.5 Tools for the characterisation of the cell wall components

The composition and organisation of the plant cell wall can be investigated by combining deconstructive and holistic approaches. These approaches consist of the extraction of specific components from cell wall materials, and their biochemical characterisation. In this manner, information about the linkages between the monomeric units and their arrangement along the polymers can be obtained and used for the construction of cell wall models by predicting their physicochemical behaviour in the wall (Busse-Wicher et al., 2014). The spatial architecture of cell wall components in the cellular contexts can also be investigated by the use of specific molecular probes on plant materials and visualisation techniques (Lee et al., 2011). The comprehension of particular structural contexts of the cell wall can then be achieved by using a combination of bioinformatics, molecular techniques and biochemical assays.

Sometimes several approaches are required to fully characterise specific aspects of the cell wall, and therefore, it is useful to combine different approaches that can provide similar sort of information. To determine the dynamics of the cell wall components it is crucial to describe their detailed structure and their biosynthesis mechanisms, and then merge the information provided between the two (Lee et al., 2011). The structural characterisation of cell wall components usually involves their extraction, treatment with chemical or biochemical agents, separation and/or purification and detection. A brief description of some of these processes are described in this section for the study of xylan, since it was the main cell wall component investigated for this thesis.

1.5.1 Xylan extraction from the cell wall

Different methods exist to extract xylan, which vary on the extent to which its different structural aspects are preserved (Hägglund et al., 1956). Extraction methods using alkali conditions generate relatively high yields of xylan, and remove all ester bonds including the ones present in the Ac modifications on the xylan backbone. A less efficient method using dimethyl sulfoxide (DMSO) has been described and used for the extraction of xylan which preserves Ac modifications on the backbone (Gonçalves et al., 2008).

1.5.2 Enzymatic hydrolysis

The use of glycosylhydrolases (GHs) with known substrate specificity to digest the cell wall polysaccharides can provide information about the structure of these polysaccharides (Fry, 2011). The GHs binding sites and substrate cleavage points of have been categorized according to the nomenclature proposed by Biely et al. (Biely et al., 1997). In this nomenclature, the substrate is drawn with the RE at the right-hand side (Figure 1.28). The glycosidic bond that is hydrolysed (cleavage point) is considered the reference point. The substrate backbone residues (subsites) are numbered according to their relative position to the cleavage point.

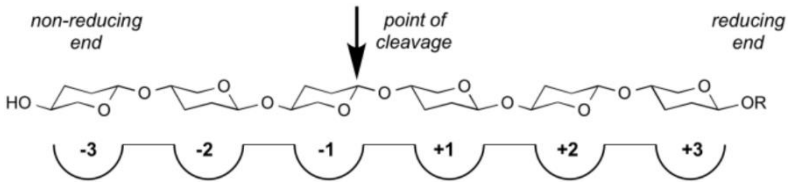


Figure 1.28: Diagrammatic representation of the sugar binding sites in glycosylhydrolases.
The substrate is drawn with the RE on the right-hand side, and the backbone residues are numbered according to their relative distance to the enzymatic cleavage point.
Taken from CAZYPEDIA (Davies et al., 1997).

GHs are classified in families in the CAZy database (www.cazy.org) according to the reaction type that they catalyse, reaction mechanism, whether they act on the ends or the middle of the substrate (exo/endo), and sequence similarity (Cantarel et al., 2009). The capability of accommodating xylan substitutions in the different subsites varies between GH families. For example, GH11 are endo-β-1,4-xylanases that can accommodate backbone Xyl substituted with 2-*O*-[Me]GlcA, one or two *O*-acetyl groups, or 3-*O*-Ara*f* at the -3 and +2 subsites, but cannot tolerate these substitutions when they are closer to the cleavage site (Chapter 6, Biely et al., 1997; Chiniquy et al., 2012; Vardakou et al., 2008). This means that GH11 requires three consecutive undecorated backbone residues to hydrolyse the xylan backbone (Figure 1.29). In contrast, GH10 endo-β-1,4-xylanases can tolerate decorations at the -2 and +1 subsites, which is a property that enables GH10 to digest more heavily-substituted xylan (e.g. substituted with the main alternating pattern of acetylation, Figure 1.21) (Biely et al., 1997; Vardakou et al., 2008).

Other endo-β-1,4-xylanases require specific substitutions on xylan to bind to the substrate. GH30 glucuronoxylanases require a backbone Xyl carrying a 2-*O*-[Me]GlcA and/or Ac at -2 subsite, and an unsubstituted Xyl at -1 (Bromley et al., 2013; Busse-Wicher et al., 2014; St John et al., 2011). This substrate specificity makes GH30 a remarkable tool for detecting [Me]GlcA on xylan, and for determining the spacing between them (Figure 1.31 panel A).

Exo-xylanases are also useful for the analysis of xylan. GH3 β-1,4-xylosidases can hydrolyse unsubstituted backbone Xyl from the NRE when the O-3 of the Xyl at +1 is unsubstituted (Tenkanen et al., 1996). GH3 can be used to detect the presence of unsubstituted Xyl on xylooligosaccharides, and to identify the existence of sidechains on the backbone of xylooligosaccharides, since their occurrence obstructs GH3 activity (Figure 1.30). This GH3 feature allows differentiating between isobaric structures when analysed by mass-spectrometry (Figure 1.31 panel B).

Other GH families such as GH115 and GH62 have also been characterised as xylan-specific but their activity hydrolyses linkages of sidechain modifications instead of links between backbone residues. GH115 glucuronidases are capable of removing O-2-linked GlcA and MeGlcA substitutions on the backbone Xyl (Biely et al., 2000). The activity of *Talaromyces* GH115 is hindered by Ara*p* modifications on GlcA (Mortimer et al., 2015), which makes this enzyme a useful tool for identifying whether GlcA substitutions form part of a PUX structure (Figure 1.30). GH62 has two hydrolytic activities; It is capable of cleaving 1,2 and 1,3-linked single Ara*f* substitutions on the backbone (Kellett et al., 1990).

In addition to GH, acetylesterases can be useful for the analysis of acetylated polysaccharides. Xylan-specific acetylesterases have been described, capable of removing Ac with different regiospecificities (Puchart and Biely, 2015). In this thesis a xylan acetylesterase from the CE4 family was used, which specifically removes Ac from mono-substituted Xyl (Figure 1.30). Modifications on di-substituted Xyl are insensitive to CE4 activity. The use of CE4 in combination with GH11 xylanase can provide information of neighbouring di-substituted Xyl on xylan (Figure 1.31 panel C).

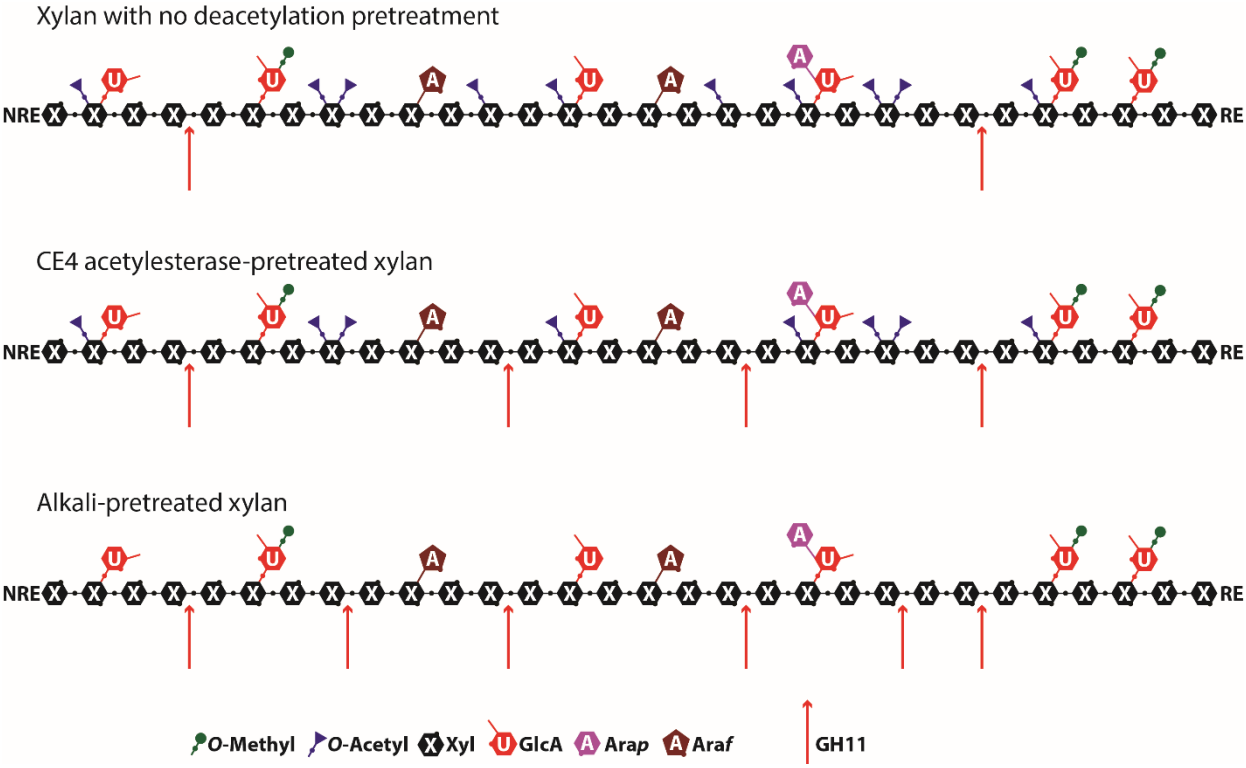


Figure 1.29: Impact of acetylation of xylan on enzymatic hydrolyses with GH11 endoxylanase.
GH11 requires three consecutive unsubstituted Xyl to perform a cleavage. Intact xylan in eudicots is mostly acetylated on every other Xyl and only few GH11 cleavage sites are available. Treating the xylan with CE4 acetylesterase removes Ac from mono-substituted Xyl of the backbone, which greatly increases the accessibility to the backbone by GH11. The pretreatment or extraction of xylan with alkali removes all Ac from the backbone, allowing higher accessibility by GH11 to the backbone than the treatment with CE4.

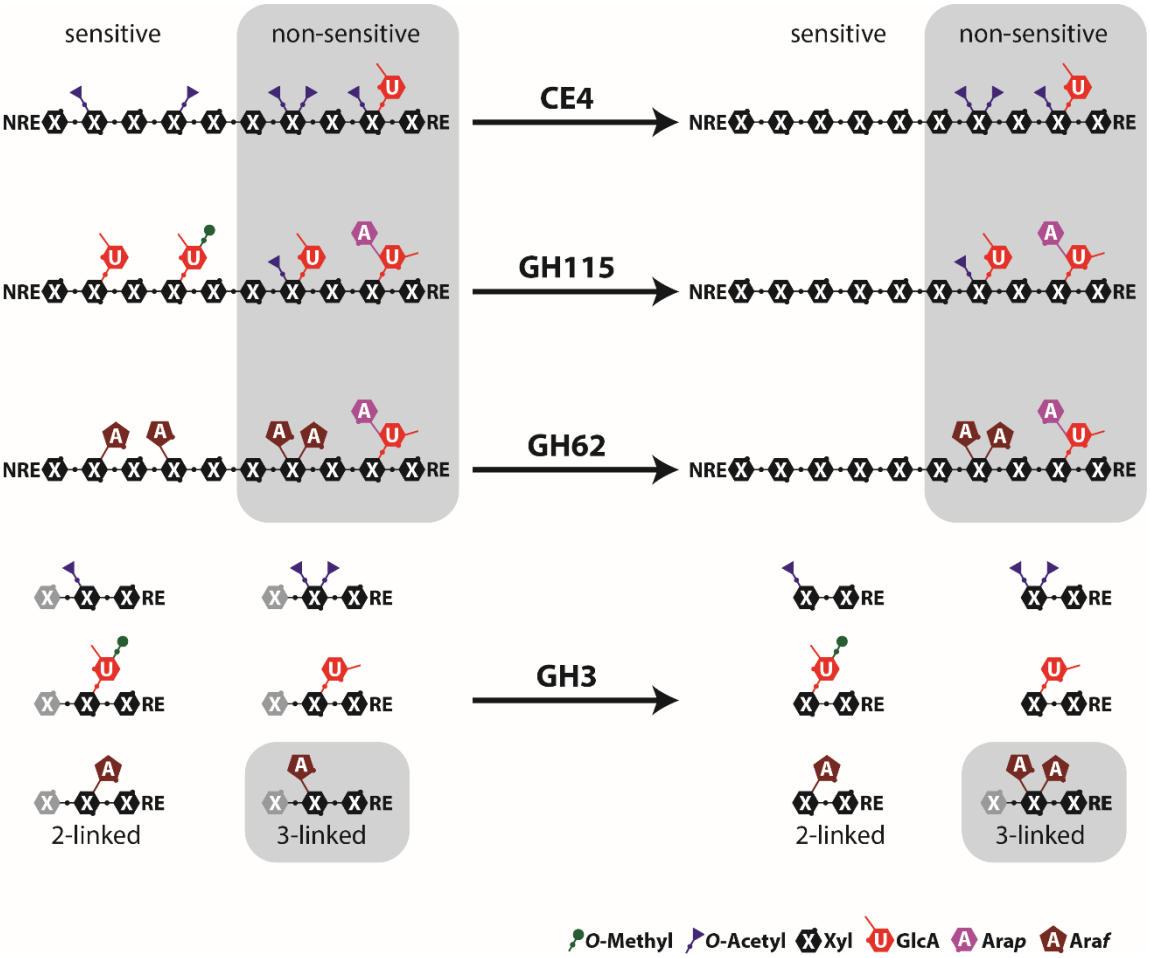


Figure 1.30: Xylan-active enzymes for analysis of xylooligosaccharides.
CE4 acetylesterase removes Ac from mono-substituted Xyl of the backbone. GH115 glucuronidase hydrolyses 2-*O*-GlcA and 2-*O*-MeGlcA from mono-substituted Xyl. GH62 arabinofuranosidase hydrolyses 2-linked and 3-linked Ara*f* from mono-substituted Xyl. GH3 β-1,4-xylosidase removes unsubstituted backbone Xyl from the NRE. GH3 activity is inhibited by 3-*O*-Ara*f* substitutions in the +1 subsite. Insensitive structures are highlighted in grey.

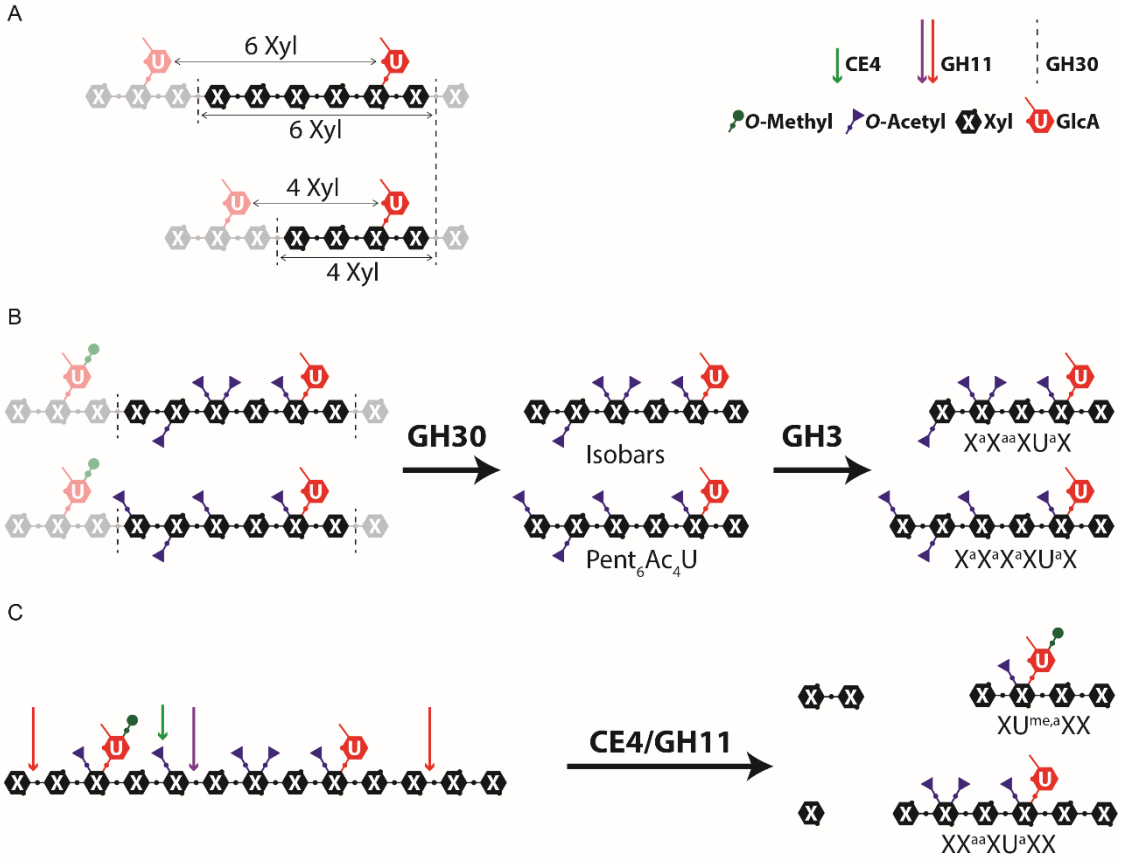


Figure 1.31: Main combined enzymatic digestions used in this thesis for xylan analysis. Panel A shows a scheme of GH30 glucuronoxylanase hydrolysis of xylan to determine the spacing between GlcA substitutions. GH30 hydrolyses the backbone with a [Me]GlcA at -2 subsite, and the products carry a single [Me]GlcA substitution at the second backbone Xyl from RE. Panel B shows a sequential digestion of xylan for determining the backbone substitution pattern between [Me]GlcA. GH30 action on acetylated xylan does not tolerate substitutions on the -1 subsite. The action of GH3 eliminates isobaric in the mix by hydrolysing undecorated Xyl from the NRE. Panel C shows a simultaneous incubation of an example xylan molecule for analysis on PACE. CE4 activity allows GH11 to gain access to the middle region of the oligosaccharide (purple arrow). On the other hand, GH11 cannot access between neighbouring substituted Xyl that are at one or two backbone residues of distance.

1.5.3 Separation and purification of oligosaccharides

Enzymatic hydrolysis of xylan from cell wall preparations can generate numerous different products. These products vary in mass, length (of the backbone), number of sidechains, etc; and they contain information of the original polymer. To characterise these oligosaccharides, it is useful to separate or purify them from the hydrolysis product mixtures.

Chromatographic methods are most commonly used for the purification and separation of oligosaccharides (Fry, 2011). These methods are based on the use of two phases, one stationary (solid) and one mobile (liquid or gaseous). In chromatographic separations the mobile phase is passed over the stationary phase (Harvey, 2000). The sample is injected into the system, and separation is achieved by a differential interaction between the oligosaccharides and the solid phase. Those oligosaccharides who form stronger interactions with the stationary phase will require longer timer to pass through the system. These interactions can be driven by different properties such as size (e.g. size-exclusion chromatography), charge (anion or cation-exchange) or polarity (normal-phase or reverse-phase chromatography). Size-exclusion chromatography separates small molecules from larger ones by delaying small molecules through interactions with the pores of smaller sizes in the stationary phase. Anion and cation-exchange chromatography retains negatively and positively-charged substrates respectively, by affinity interactions with opposite electrical charges. Polarity-based chromatography uses hydrophilic stationary phases to bind hydrophilic molecules (e.g. hydrophilic-interaction liquid chromatography, HILIC), or hydrophobic stationary phases to bind hydrophobic molecules (e.g. octadecyl-carbon chain (C18)-bonded silica columns). Then the retained groups are eluted by using mobile phases with similar polarity (Neue and Phoebe, 1997).

1.5.4 Detection and characterisation of oligosaccharides

Separation and detection methods can be coupled together. Polysaccharide Analysis by Carbohydrate gel Electrophoresis (PACE) couples electrophoretic separation to the highly sensitive detection of fluorophore-labelled oligosaccharides (Figure 1.32) (Goubet et al., 2002). The fluorophore is linked to the RE of the sample oligosaccharides by reductive amination, and fluorescence is detected on

polyacrylamide gels (PACE). The charged oligosaccharides are exposed to an electric field, and their migration velocity will depend on the molecular size and the total molecular charge of the oligosaccharide. In this way, the labelled digestion products produce bands on PACE with characteristic mobility. Enzymatic profiles can be annotated by using further enzymatic digestions, mass-spectrometry and NMR methods. Once an enzymatic profile for the digestion products of a specific GH combination is annotated, this profile can be used as reference for the assignment of structures in other samples. This oligosaccharide characterisation method is based on the assumption that digestion products that exhibit the same electrophoretic mobility correspond to similar structures. Therefore, PACE is a fast and simple method for describing structures of xylan and other wall polysaccharides, and was thoroughly used in the work carried out for this thesis.

For the case of acetylated oligosaccharides, separation by electrophoresis may be challenging because of the spontaneous migration of Ac within sugar residues (Kabel et al., 2003; Figure 1.15). During separation methods, xylooligosaccharides carrying Ac on mono-substituted Xyl exhibit a dispersed distribution and do not separate (Busse-Wicher et al., 2014). It is believed that this dispersion is caused by the change in velocity through the system by the isomeric oligosaccharides attributed to the specific position of the Ac, and also by the change in Ac position during electrophoretic resolution. This is translated into the production of smeary bands on PACE.

To fully describe the structural aspects of acetylated and non-acetylated polysaccharides and oligosaccharides it is necessary to use mass-spectrometric and NMR techniques. Tandem mass-spectrometry can accurately provide information of the position of the modifications on the backbone of xylooligosaccharides (Busse-Wicher et al., 2016b, 2014; Chong et al., 2014; Mortimer et al., 2015). On the other hand, 2-D solution-state NMR experiments can provide a quantitative outcome for the occurrences of specific structures in the xylan polymer (Busse-Wicher et al., 2014; Xiong et al., 2013).

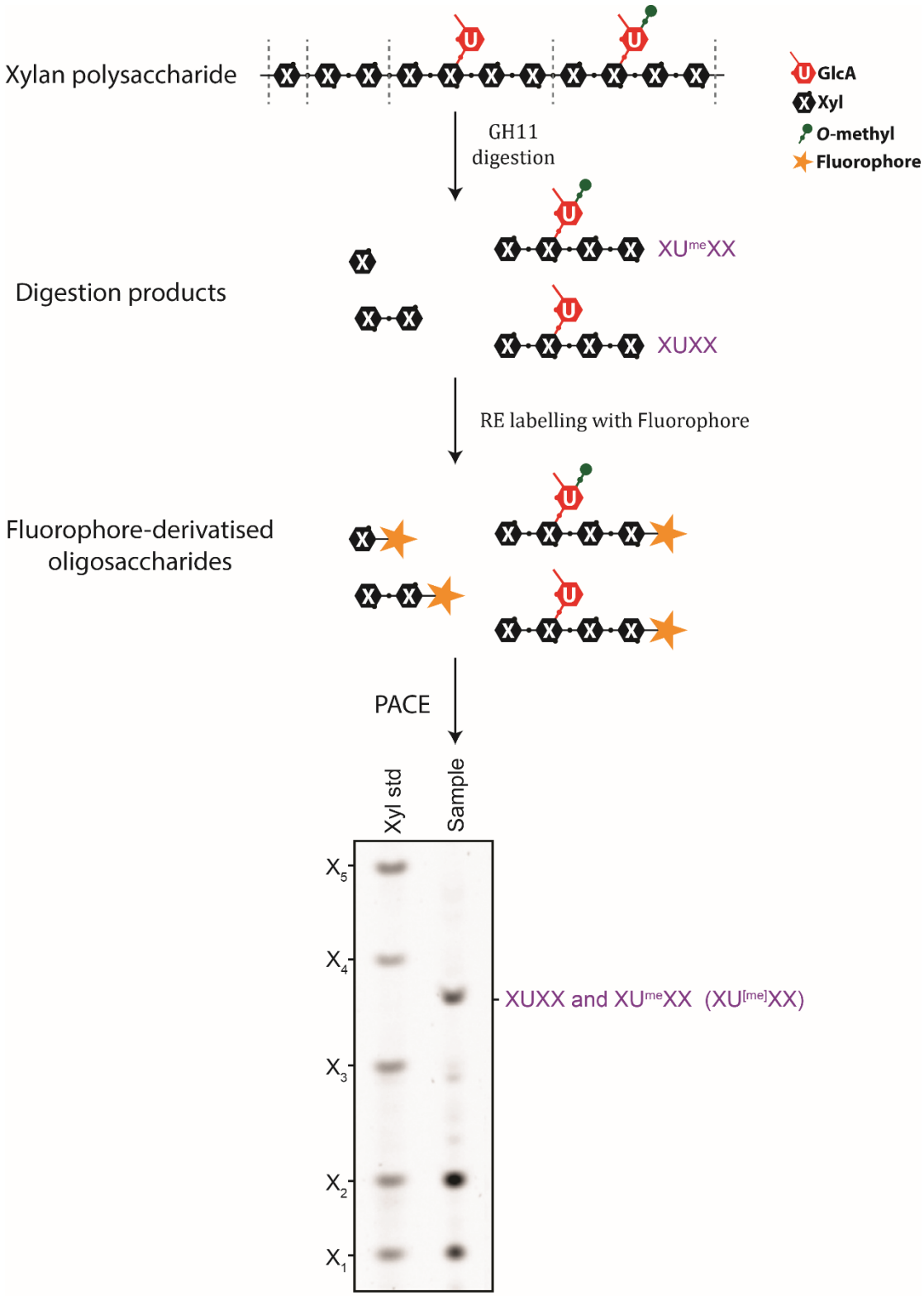


Figure 1.32: Workflow diagram for PACE and DASH analyses using GH11. Deacetylated xylan from *Arabidopsis* bottom stems is shown as substrate which is digested with GH11. The digestion products are xylose (X_1), xylobiose (X_2), GlcA-substituted xylotetraose (XUXX) and MeGlcA-substituted xylotetraose ($XU^{me}XX$). DASH higher resolution is revealed by the separation of XUXX and $XU^{me}XX$. Xyl std: Xylooligosaccharide standards X_1-X_5. Fluorophore used: ANTS. PACE gel image is reproduced from Mortimer et al. (2010).

1.6 Project aims

Xylan structure and biosynthesis are not fully understood, yet they are both important for understanding how plant cell walls are structured. This project consisted of the detection, description and characterisation of different acetylated xylan structures, and the identification of different enzymes involved in the biosynthesis of acetylated glucuronoxylan. The objectives of this thesis are divided in the following sub-aims, which were approached in the chapters of this thesis.

1) The acetylation on xylan by ESK1 determines the GlcA pattern (Grantham et al., 2017), and this [Me]GlcA pattern seems altered in secondary cell wall XSC mutants. In the absence of IRX9, IRX10 or IRX14, the respective homologue proteins can synthesise xylan (Wu et al., 2010, 2009), but perhaps the acetylation pattern is different causing a disruption of the [Me]GlcA pattern. The first sub-aim was to analyse if there is a mechanistic link between the secondary cell wall XSC and ESK1, and whether this link is affected by altering the composition of the XSC. In order to do this, the xylan acetylation pattern was analysed in mutants with different XSC compositions.

2) Mostly, acetylated Xyl and undecorated Xyl alternate on the xylan backbone (Busse-Wicher et al., 2014; Chong et al., 2014). However, acetylation in all regions of secondary-cell-wall xylan have not been previously described. Therefore, the second aim of this project was to analyse and distinguish the acetylation patterns in the compatible and in the incompatible domains of xylan. In order to do this, the areas of xylan surrounding of [Me]GlcA decorations were characterised in *gux1* and *gux2 Arabidopsis* mutants.

3) The presence of acetylation on primary-cell-wall xylan could be inferred by combining the results obtains in two independent publications (Chong et al., 2014; Mortimer et al., 2015), yet no direct evidence has been shown. The third aim was to determine if primary-cell-wall xylan from different tissues is acetylated, and to describe the substitution patterns in the acetylated xylans.

4) Xyl on xylan can be mono-acetylated on O-2 or O-3, can be di-acetylated on O-2 and O-3, and can also be substituted with a [Me]GlcA on O-2 and/or an Ac on O-3. ESK1 has been characterised as a xylan-specific O-2 acetyltransferase (Urbanowicz et al., 2014). However, the proteins responsible for the Ac in other positions were not described before. The last aim of this thesis was to identify additional acetyltransferases acting on xylan, and to describe the specific positions of the backbone Xyl they act on.

Chapter 2: Materials and Methods

2.1 *In silico* work and bioinformatics

2.1.1 Phylogenetic trees

2.1.1.1 Sequences acquisition

The *Arabidopsis thaliana* (AT) amino acid sequences of ESK1 and the whole TBL family were obtained from TAIR (arabidopsis.org), with a total of 46 sequences. ESK1 homologue sequences in other species were found using the BLAST tool on the website https://bioinformatics.psb.ugent.be/plaza by using as input ESK1 amino acid sequence. Sequences from the following species were acquired: *Amborella trichopoda* (ATR), *Brachypodium distachyon* (BD), *Cycas micholitzii* (CMI), *Eucalyptus grandis* (Eucgr.), *Ginkgo biloba* (BGI), *Gnetum montanum* (GMO), *Oryza sativa* (OS), *Picea glauca* (PGL), *Physcomitrella patens* (PP), and *Taxus baccata* (TBA) were acquired from the online platform PLAZA (Van Bel et al., 2012), version Gymno v1.0 (updated in June 2018) with a total of 429 including the sequences from *Arabidopsis*.

2.1.1.2 Phylogenetic tree construction

Three trees were constructed in this thesis (in order of appearance):

1) The tree in Figure 6.1 (page 177) was constructed with the *Arabidopsis* sequences only. The sequences were aligned using the MUSCLE algorithm. A maximum likelihood (ML) phylogenetic tree was constructed using MEGA7 software following the default instructions published by Hall (2013), and applying a bootstrap (*n* = 1000 trials). The alignment file used to construct the tree is shown in the annex (Figure S3).

2) The tree shown in Figure 6.3 (page 181) was constructed using the sequences from *Arabidopsis thaliana* (AT), *Amborella trichopoda* (ATR), *Brachypodium distachyon* (BD), *Eucalyptus grandis* (Eucgr.) and *Oryza sativa* (OS). The full-length sequences were aligned using the MUSCLE algorithm, and then the variable N-terminal region was removed. A ML tree was constructed following the default instruction published by Hall (2013). The best substitution model

found was LG model with γ distributed and invariant sites. These parameters were used to build a maximum likelihood tree with 1000 iterations bootstrap using the MEGA7 software. The alignment file used to construct the tree is shown in the annex (Figure S4).

3) The tree in the annex (Figure S5) was constructed using the full amino acid sequences from all species mentioned in section 2.1.1.1. The sequences were aligned using the MUSCLE algorithm and then highly variable, gapped N-terminal regions were removed. A ML tree was constructed using FastTree software with a Shimodaira-Hasegawa robustness test. The alignment and the tree are shown in the annex (Figure S6).

2.1.2 Gene expression data

Gene expression values for *GUX1*, *GUX2*, *GUX3, TBL3*, *TBL28*, *TBL30*, *TBL31*, *TBL32*, *TBL33*, *TBL34*, *TBL35* and *TBL36*, were extracted from three different sources, and the numeric values were managed using Microsoft Excel 2016:

1) RNA microarray data was obtained from the eFP Browser website, specifically from the *Arabidopsis thaliana* Tiling Array Express database (At-TAX, http://gbrowse.weigelworld.org/cgi-bin/attax) (Laubinger et al., 2008). Only absolute values were downloaded. For each gene, the mean (K) of the values from different tissues was calculated, then each absolute value was divided by K (mean normalised).

2) RNAseq data was acquired from TRAVA database (http://travadb.org/) (Klepikova et al., 2016), using the “Raw Norm” option when browsing to obtain absolute values. Transcript copy number values per tissue were used without further modifications.

3) ATH1 microarray expression data (Waese et al., 2017) was available only for *GUX1*, *GUX2*, *GUX3*, *TBL3*, *TBL28*, *TBL30*, *TBL32* and *TBL34*. Plant developmental stages and seed-specific expression pattern diagrams were downloaded from the eFP Browser website (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) (Winter et al., 2007).

2.1.3 Thesis composition

All word processing and writing was carried out using Microsoft Word 2016 (Microsoft, Redmond, WA, US). Cites and references were processed and formatted using the software Mendeley Desktop version 1.19.2 (Mendeley Ltd., Elsevier, Amsterdam, Netherlands).

Numeric data processing, tables and graphs were carried out using Microsoft Excel 2016 (Microsoft, Redmond, WA). DASH data was processed using the software DASHBOARD (Li et al., 2013). Mass spectrometric data was initially processed using the software FlexAnalysis version 3.4 build 76 (Bruker, Billerica, MA, US), and then formatted with Microsoft PowerPoint (Microsoft, Redmond, WA).

Image processing, contrast and brightness were adjusted using Adobe Photoshop CS6. Further figure preparations, backgrounds, labels and other vector-graphic modifications were carried out using Adobe Illustrator CS6 (Adobe, San Jose, CA, USA).

2.2 Biological material

2.2.1 Bacteria

2.2.1.1 *E. coli* strain and transformation

Escherichia coli strain DH5α (C2988J, New England Biolabs) was made chemocompetent and transformed following the protocol published by Sambrook et al. (2006). The growth medium used was LB (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract) supplemented with antibiotics and using 2% agar in solid cultures. The antibiotics used for transformants selection varied depending on the plasmid transformed. The antibiotics used for the selection of each plasmid type are listed in Table 2.1.

Table 2.1: List of antibiotics used for the selection of each plasmid type.

Antibiotic	Concentration (mg/L)	Plasmid selection
Spectinomycin (Spec)	50	L0 modules
Ampicillin (Amp)	100	L1 transcriptional units
Kanamycin (Kan)	50	L2 binary vectors and *pGreenII0000*
Gentamycin (Gen)	50	Ti helper plasmid

2.2.1.2 *A. tumefaciens* strain and transformation

Agrobacterium tumefaciens strain *GV3101* was used, which has a rifampicin (Rif) resistance gene integrated in its genome, and a Ti helper plasmid with a Gen resistance gene selection marker.

To prepare chemocompetent *A. tumefaciens*, a 5 mL LB medium (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract) supplemented with Gen (50 mg/L) and Rif (10 mg/L) was inoculated with a single colony. The culture was grown overnight, and 3 mL were used to inoculate 100 mL LB medium supplemented with Gen and Rif. The culture was incubated in a 250 mL conical flask shaking at 250 rpm for about 3 h at 30 °C until OD_{600} between 0.5 and 1.0, read with a Spectronic Helios Gamma spectrophotometer (Thermo). Then the culture was chilled on ice and pelleted at 3200 RCF for 5 min. The supernatant was discarded, and the pellet was resuspended in 2 mL of ice-cold sterile 20 mM $CaCl_2$ + 10% (v/v) glycerol. The suspension was divided into 100 µL aliquots of chemocompetent cells. The aliquots were then snap-frozen in liquid nitrogen and stored at -80 °C until use.

To transform *A. tumefaciens* one aliquot of chemocompetent cells was thawed on ice and ~1 µg of plasmid DNA was added. Then the cells were snap-frozen and incubated in liquid nitrogen for 5 min and thawed at 37 °C. The cells were supplemented with 250 µL of LB medium and incubated for 3 h at 30 °C. The cells were then plated on agar-solidified LB (2%) containing Rif (10 mg/L), Gen (50 mg/L) and Kan (100 mg/L). The plates were incubated at 30 °C, and transformed colonies were obtained after 48 h of incubation.

2.2.2 *Arabidopsis thaliana*

Almost all *Arabidopsis thaliana* plants used in this work were of the ecotype *Columbia-0* (*Col-0*) with mutant lines compared to a corresponding WT control. The only exception was *tbl33-2*, which is of ecotype *Landsberg erecta-0* (*Ler-0*).

The list of mutants that were acquired from other researchers is shown in Table 2.2. In addition, the insertional mutants for the TBL genes used in this work are listed in Table 6.1 (Chapter 6, page 198).

Table 2.2: *Arabidopsis* lines obtained from elsewhere used in this study.

Mutant name	References	Provided by
irx9	(Wu et al., 2010)	Mathias Sorieul
irx10	(Brown et al., 2009; Wu et al., 2009)	Mathias Sorieul
irx14	(Wu et al., 2010)	Mathias Sorieul
irx15/15-l	(Brown et al., 2011; Jensen et al., 2011)	Xiaolan Yu
gux1	(Bromley et al., 2013; Mortimer et al., 2010)	Mathias Sorieul
gux2	(Bromley et al., 2013; Mortimer et al., 2010)	Mathias Sorieul
gux1/2	(Bromley et al., 2013; Mortimer et al., 2015, 2010)	Nicholas Grantham
gux1/2/3	(Lee et al., 2012; Mortimer et al., 2015)	Nicholas Grantham
gxm1/2/3	(Cornuault et al., 2015; Li et al., 2013)	Toshi Kotake
irx9 gux1/2	undescribed	Xiaolan Yu
irx10 gux1/2	undescribed	Xiaolan Yu
irx14 gux1/2	undescribed	Xiaolan Yu
irx15/15-l gux1/2	undescribed	Xiaolan Yu
esk1 (esk1-5)	(Grantham et al., 2017; Xiong et al., 2013)	Nicholas Grantham
axy9 (axy9-2)	(Schultink et al., 2015)	NASC
various *tbl* lines*		various*
Callus lines		
Col-0	(Mortimer et al., 2015; Nikolovski et al., 2012)	Xiaolan Yu
irx10-l	(Mortimer et al., 2015)	Xiaolan Yu
gux3	(Mortimer et al., 2015)	Xiaolan Yu

*For *tbl* T-DNA insertional lines see Table 6.1 (Chapter 6, page 198).

2.2.2.1 Seed sterilisation

Mature dry seeds were surface-sterilised by immersion for 10 minutes in a sterilising solution (80% ethanol, 1% Tween® 20 (Sigma-Aldrich, P1379) and 1% sodium hypochlorite) with vigorous shaking. Then they were successively washed with 65% ethanol, 80% ethanol and 100% ethanol. The supernatant was removed, and the seeds were air-dried in a laminar flow hood.

2.2.2.2 Growth in soil

Sterile seeds were sown on 12 x 12 cm^2 sterile plastic plates containing 25 mL of germination medium (2.2 g/L Murashige and Skoog basal medium (Sigma, M5519), 10 g/L sucrose, 6 g/L PlantAgar (Duchefa, P1001.1000), 5 g/L MES pH 6.8 adjusted with KOH). The plated seeds were incubated in the dark at 4 ºC for 48 h. Then they were incubated in long day regime (16 / 8 h light and dark photoperiod) for 14 d.

Plantlets were then transferred to pots with soil (Levington Advance Solutions which contains the insecticide Exemptor®) and grown under long day regime for 6 weeks at 21 °C. For the NMR experiments the plants were transferred to soil and grown in short day regime (8 / 16 h light and dark photoperiod) for 4 weeks and then grown in long day regime for 4 extra weeks. After growth, the plants were left to dry by not watering them for two weeks.

2.2.2.3 Growth in magenta pots

Due to the nature of the mutants *axy9* and *esk1 tbl33* being extremely sick and dwarfed, they were first grown for 14 days in plates with germination medium, and then they were transferred to glass magenta pots containing 100 mL of germination medium. Three plants per pot were grown for 6 additional weeks.

2.2.2.4 Mature stem tissue collection

About a fifth of the primary stems were harvested. The material was stored at room temperature until used. Prior to AIR preparation they were immersed in 70% ethanol and incubated at 70 °C for 30 min.

2.2.2.5 Young stem tissue collection

The collection of young stem material was carried out as described by Mortimer et al. (2015). The collected stems were incubated in 96% ethanol at 70 °C for 30 min. The material was then stored at 4 °C until used.

2.2.2.6 Callus growth and collection

Callus lines were grown by Xiaolan Yu following the method described by Mortimer et al. (2015). The collected calli were washed twice with 96% ethanol. Then they were incubated in 70% ethanol for 30 min. The material was then stored at 4 °C until used.

2.2.3 Birch

The birch samples were kindly provided by Maija Tenkanen and Ykä Helariutta (University of Helsinki, Finland). Three biological independent birch trees (*Betula pendula*, v5834) were grown in the experimental field located at the Viikki campus of the University of Helsinki. Woody stems from each were collected for cryo-sectioning to harvest tangential section of tissue-specific wood fractions. Young phloem and young xylem were dissected through a cryotome (thickness: 10 µm, -27 °C) and separately collected. The tissue samples were freeze-dried, and ground using a bead mill until fine powder was obtained.

2.3 Molecular biology

2.3.1 *Arabidopsis* mutants genotyping and expression analysis

2.3.1.1 DNA extraction

Seedlings were grown for 10 days. One cotyledon was extracted and ground in 100 µL of TNE-SDS buffer (200 mM Tris-HCl, 250 mM NaCl, 20 mM Ethylenediaminetetraacetic acid (EDTA), 0.5% SDS; pH 8.0). The mixtures were centrifuged at 3200 RCF for 5 min. The pellets were discarded, and the supernatants were transferred into a new tube. 100 µL of isopropanol were added to the samples and mixed by vortexing. The samples were then incubated for 30 min at -80 °C and the precipitated DNA was sedimented by centrifuging at 3200 RCF for 10 min. The supernatants were discarded, and the pellets were washed twice with 70% ethanol. The remaining ethanol was evaporated from the samples by exposing them to vacuum for 10 min. The DNA samples were then resuspended in 30 µL of H_2O.

2.3.1.2 Primers and PCR parameters for screening T-DNA insertions

The oligonucleotide primers for genotyping *Arabidopsis* lines were individually designed for the analysis of each T-DNA insertion. The primers were designed using the online tool iSectPrimer toolbox (http://signal.salk.edu/tdnaprimers.2.html) using the default parameters and synthesised by Sigma-Aldrich (UK). The primers used for

screening insertions in the TBL genes are listed in Table 6.1 (Chapter 6, page 198), and all the rest are listed in Table 2.3. Right primer (RP) and left primer (LP) combined were used for screening wild-type gene versions, and for mutant alleles RP and border primer (BP) were used. To genotype T-DNA lines the REDTaq ReadyMix PCR Reaction Mix (Sigma, R2523) was used according to the manufacturer's instructions. Each reaction contained 8 µL of RedTaq DNA polymerase, 0.25 µL of each primer (final concentration 4.8 µM), 0.25 µL of DMSO, 2.75 µL of H_2O and 0.5 µL of extracted DNA. The mixture was then incubated in a Veriti Thermal Cycler (Applied Biosystems) following the program shown in Table 2.4. PCR products were analysed on an agarose gel, purified from the gel and sequenced.

The mutants *irx9 gux1/2*, *irx14 gux1/2*, were initially crossed by Xiaolan Yu, and *irx10 gux1/2* by Marta Busse-Wicher, and the genotyping was carried out in the work for this thesis. *irx15/15-l gux1/2* was crossed and genotyped by Xiaolan Yu.

Table 2.3: List of gene mutations and primers used for their screening.
RP Right primer; LP Left primer; BP border primer.

Mutant gene	**Gene ID**	**Insertion**	**PCR primer codes**		
			RP	**LP**	**BP**
irx9	*At2g37090*	SALK_057033	1360	412	248
irx10	*At1g27440*	SALK_046368	748	744	249
irx14	*At4g36890*	SALK_038212	500	501	248
gux1	*At3g18660*	SALK_063763	1016	1015	249
gux2	*At4g33330*	GABI_722F09	1014	1013	476

Table 2.4: PCR program.

Step	**Temperature (°C)**	**Time**	**Cycles**
1	95	4 min	1
2	95 58 (-0.5 °C every cycle) 72	30 s 30 s 2 min	10
3	95 58 72	30 s 30 s 2 min	35
4	72 4	4 min -	1

2.3.1.3 RNA Extraction

Total RNA extractions from stems and leaves were carried out with Spectrum Plant Total RNA Kit (Sigma-Aldrich, STRN50), following the manufacturer's instructions. One rosette leaf or one stem of a 6-week old plant was extracted, and snap-frozen in liquid nitrogen and used for the RNA extraction procedure. The RNA quality and quantity were tested by using a Nanodrop Spectrophotometer (Thermo Scientific, ND-1000).

2.3.1.4 RT-PCR

RT-PCR was used to test the presence of transcript of *TBL3*, *TBL28*, *TBL30*, *TBL32*, *TBL33*, *TBL35* and *TBL36* in leaves and stems, in WT and the respective T-DNA insertional lines. Extracted RNA was used for production of complementary DNA (cDNA) by using SuperScriptIII First-Strand Synthesis System for RT-PCR (Invitrogen, 18080-051). The procedure was carried out following the Oligo$(dT)_{20}$ version of the protocol. The quality of the cDNA obtained was verified with a Nanodrop Spectrophotometer (Thermo Scientific, ND-1000) and used as template for PCR. The primers used for the PCR are listed in Table 6.1 (Chapter 6, page 198). In addition, *HISTONE1* was amplified using primers 510 and 511, and used to normalise the amount of template for each reaction. The PCR program used is a modified version of the one described in Table 2.4, without step 2, and the time of the 72 °C in step 3 varied according to the RedTaq manufacturer's instructions (Sigma-Aldrich, R2523) depending on the length of the amplification products. The PCR products were analysed by migration on an agarose electrophoresis gel, and by sequencing (as in section 2.3.1.2).

2.3.1.5 Electrophoresis analysis and DNA purification on agarose gels

PCR products or DNA digestion products were electrophoresed at 200 V on 1.2 % agarose gels containing 0.4% ethidium bromide (Sigma-Aldrich). Hyperladder 1kb (Thermo Fisher, 10787018) or 50bp DNA Ladder RTU (GeneDireX, DM012-R500) DNA mobility markers were used depending on the sample expected molecular size. Products were visualized using the U:Genius3 gel imaging system (Syngene).

For DNA purification and recovery from the agarose gel, the band was excised with a knife. Then the DNA was purified using Wizard SV gel and PCR cleanup system (Promega, A9281), following the manufacturer's instructions. The quality and amount of DNA was tested using a Nanodrop Spectrophotometer (Thermo Scientific, ND-1000).

2.3.1.6 DNA sequencing

PCR or RT-PCR products were directly purified using Wizard SV gel and PCR cleanup system (Promega, A9281), following the manufacturer's instructions. Alternatively they were resolved on an agarose gel and then extracted and purified (see above, section 2.3.1.5). The purified DNA samples were sequenced by the DNA sequencing facility in the Department of Biochemistry of the University of Cambridge, with both primers previously used for the PCR/RT-PCR amplification.

2.3.1.7 Crossing

Several *tbl* T-DNA insertional lines were cross-pollinated, and the lines obtained are listed in Table 2.5. To cross *Arabidopsis* lines, the procedure described by Weigel and Glazebrook (2006) was followed.

Table 2.5: Lines obtained by crossing.

Cross	lab code
esk1 tbl28-1	BMQ
esk1 tbl28-2	BMR
esk1 tbl30-1	BMS
esk1 tbl30-2	BMT
esk1 tbl3-2	BMP
esk1 tbl32-1	BMU
esk1 tbl33-1	BMV
esk1 tbl33-2	BMW
esk1 tbl35-1	BNV
gux1 tbl33-1	BGV
gux1 tbl33-2	BGX
gux2 tbl33-1	BGW
gux2 tbl33-2	BGY
tbl32-1 tbl33-1	BMY
tbl32-1 tbl33-2	BMZ

Table 2.6: Codes and sequences of DNA oligonucleotides used for molecular biology.

Primer code	Sequence (5' - 3')
210	*CATTTTATAATAACGCTGCGGACATCTAC*
248	*AACCAGCGTGGACCGCTTGCTG*
249	*TTTTTCGCCCTTTGACGTTGGAG*
396	*CCGGATCGTATCGGTTTTCG*
405	*GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC*
412	*ACTCACTCAAGAAACCTTCTTCC*
476	*ATATTGACCATCATACTCATTGC*
500	*AACGACACGTGTACCTCCTTG*
501	*TGCCTCAAAGCTTGAAGTCTC*
510	*GGCTGCTCCGGTTAAGAAGAA*
511	*TTTGGAAAGCAAGCTCCTTCA*
744	*GGAAAAAGCCATTGAAAGAGG*
747	*TGGTTCACGTAGTGGGCCATCG*
748	*CCACTCGGAGGACTTGGA*
1013	*GTGAAATTGATGTGAAAGATACCAAC*
1014	*AAAGTGCTAATTAATCACCCTTGATC*
1015	*ACACAACTGATATTCACCTAATCTCG*
1016	*TAATCTTGAGAAACATCGATTTCTTG*
1234	*CTTCGTCGTAGTCTTCTCTGTTTTC*
1235	*ACTTACAGGACTTTCATAGCGAATG*
1298	*AACCTAGTAAGCACATACGCGATAC*
1299	*CTCTCTCCATCATCGTTCTCG*
1301	*TGTTGACCAAAGGAAAAATAGGTAG*
1302	*GCTTATTTTTGTTCTCAGAAAGCTG*
1303	*GCTTATTTTTGTTCTCAGAAAGCTG*
1304	*GCTTATTTTTGTTCTCAGAAAGCTG*
1305	*GTTGAAGACTATGGTTAAATGGGTG*
1309	*AACATGGTCGTCACTGGAAAG*
1310	*ACAACTTTTATGCCTTGAGTTTGAC*
1312	*ATGCTAGATATTGGCGTGGTG*
1313	*TAACTTCAAATCCCTTGAATCTCTG*
1314	*CTTGGCTCTTTAACTCACTGTTCTC*
1315	*AACCTAGTAAGCACATACGCGATAC*
1316	*CTGTAAGCCACTTGTGTGTCTAGTG*
1317	*TATCACGCTTGACACTACTTTCATC*
1318	*CTCTCCATTTCCGAAAGAACC*
1319	*AATTTGGTAAGACTAGACCGGTTTC*
1320	*AAAAAGACCTTAAAACACACAGCAC*
1321	*TGAGGGTATAAGGTGTGCTTTAGAG*
1322	*TGTTCTGTTTTTGTTTTAGGAGCTC*
1323	*TGTTCTGTTTTTGTTTTAGGAGCTC*
1324	*TGTTAGAGTTTGAGCAGCAGTATTG*
1328	*TTTTGTCAGTTAGGTTTGGCG*

1329	*ATCACCTTCATGCTGAAGATAAATC*
1330	*AATCATCAGTTTCTACCGTTCTTTG*
1331	*ACTTAATCACCGCAATGTTCG*
1332	*GTTGTTAGAGAAGCTGAGGAACAAG*
1333	*ACTTACAGGACTTTGATGGAGACAG*
1334	*TATCACGCTTGACACTACTTTCATC*
1335	*AACCTAGTAAGCACATACGCGATAC*
1336	*CCAGAGGGCCCTAGGATGGATGTCG* *CAGAGATGGAGTAGAAAGAAGAG*
1338	*ACAATGCCACAGTGGAGTTTCTCTGGG*
1359	*ACCCGACCGGATCGTATCGGT*
1360	*AACTTACCAACCCACCCATTC*
1381	*TCATGTGCTTTATCACGCTTG*
1382	*GAAGTGCTTGACGACAGGTTC*
1383	*CATCTTTGGACAGCTCGAGCC*
1384	*GGCATGCGTTGGAGACATGC*
1564	*GAGTGTGAGTTCTTGACGGAG*
1569	*GTGAGTTTCTATCAGAGTGGGTG*
1571	*TCCGTACATTCAACCTCAGC*
1573	*GTTGGCTTGTGAGAAGTTTGG*
1575	*TCGTCAAGCCTAATGTTCCTC*
1576	*TCCTCAACACCCTCTCGTAC*
1581	*CGTCTCTTCCGGTGACTTTT*
1584	*CTTCCCTCTTAATGTCTCCAGC*
1585	*ACTTCTCCAATCACTTTCATTATGC*
1590	*ATGGCAAAGCGACAACTGTTGATGC*
1591	*CGCCTTTGGGTTTGTAATTACCG*
1604	*TCTTGTAGATTGACGTGTGC*
1605	*CTCCAAAACTCTCTTCACCA*
1606	*CACCATTGATACCCATTGGC*
1608	*CAGAGGGCCCCATTGGTAAATTACATCGTGACATTTAT*
1609	*GAGACCTAGGGTTTAAACAACCCTCTCTTCAAGGAAGAAGG*
1610	*CACAGTTTAAACCCTAGGATGGCGTCAATCCGGCG*
1611	*CAGACTGCAGTTTCATAGTTATGAGAGCCTGAAGATTCTTTTGG*
1623	*AACACCAGTGGATGCAATCTC*
1624	*GCGCAGTGTTTAAAACCTCTG*
1625	*ACAATCTGCGTTACTACTTGG*
1628	*TGGGTCAGACTTTTGCATTTC*
1629	*TTGAATCTTGTTTTGGGTTCG*
1630	*CGAATGATTACGGTCAGATTATAGACGC*
1631	*GCCAACTGGTCAGACATGAACTTGC*
1632	*GCACAACAATGGCTCTAACCTCATCAC*
1633	*CGTGAACCTCCAACCACTGAGAC*
1635	*CTTTTATACTAATGGCCATAGATAGAGAAAGGAAAG*
1636	*CCAGAACTGGAGATGGAAACC*

2.3.2 *Arabidopsis* transformation

2.3.2.1 Bacterial DNA extraction

Plasmid DNA was extracted from 5 mL liquid *E. coli* cultures grown overnight, using GenElute™ HP Plasmid Miniprep Kit (NA0150-1KT), following the manufacturer's instructions.

2.3.2.2 Restriction-enzyme-mediated vector assembly

The pGreenII-0000 vector was modified by Mathias Sorieul to express an Oleosin (AT4G25140)-eGFP fusion protein under control of the Oleosin promoter from *pFAST-G01* as a selection marker (Shimada et al., 2010). This simplified the selection of transformants, by selecting the seeds that are fluorescent when illuminated with a blue/UV light. In addition, the vector had integrated *mGFP* at 3' from an insertion site, for tagging GFP on C-terminus of the inserted gene. The insertion site had the restriction sites 5' *Apa*I and 3' *Pst*I, flanking an insert of ~2800 bp. The vector was simultaneously digested with the restriction enzymes *Apa*I and *Pst*I. The two digestion products were resolved on an agarose gel, and the one with higher molecular size (~7600 bp) corresponding to the open destination vector, was excised and purified (see section 2.3.1.5).

The 5′ UTR and promoter region of IRX9 (*At2g37090*, *PromIRX9*), 1523 bp upstream from the ATG start codon (exclusive) was amplified with PCR Master Mix with GC Buffer (New England Biolabs, M0532S) following the manufacturer's instructions, and using primers 1608 and 1609. These primers added the flanking restriction sites *Apa*I on 5', and *Pme*I followed by *Avr*II on 3', respectively. In a similar manner, the coding region of *IRX9-L* (*At1g27600*, 1796 bp) was amplified from ATG (inclusive) to the stop codon (exclusive). For this, the primers 1610 and 1611 were used, which add *Pme*I/*Avr*II in 5' and *Pst*I in 3', respectively. The PCR program used for both reactions is modified from the one shown in Table 2.4, changing the time at 72 °C to 1 min in steps 2 and 3. The amplified PCR products were purified (see section 2.3.1.5) and then digested with restriction enzymes. *PromIRX9* was digested with *Apa*I and *Avr*II, and IRX9-L coding region was digested with *Avr*II and *Pst*I. The digestion products were then resolved on an agarose gel, excised and purified (see section 2.3.1.5).

T4 DNA ligase (Thermo Scientific, EL0014) was used to fuse the open destination vector, PromIRX9 and IRX9-L coding region together. The manufacturer's instructions were followed, includin both inserts in the reaction mix in a ratio 1:1. Then the ligation product was used to transform *E. coli*, and then the vector was selected by adding kanamycin to the growth medium. The map of the produced vector is shown in Figure 3.5 (page 94). The vector was purified and used to transform *Agrobacterium*. The sequences integrated into the vector were verified by sequencing, and the correct insertion was confirmed by observing two products when digesting the vector with *Xho*I and *Apa*I, one of ~3200 bp and another of 7700 bp.

2.3.2.3 Vector assembly by Golden-Gate

For over-expressing *TBL3*, *TBL28*, *TBL29(ESK1)*, *TBL30*, *TBL31*, *TBL32*, *TBL33*, *TBL34*, *TBL35* and *TBL36* in *esk1* mutant, expression vectors were assembled using Golden-Gate MoClo technology. This technology consists of the ensemble of binary vectors (L2) from other less complex plasmids (L1) which possess a transcriptional unit. The L1 are also ensembled from less complex plasmids (L0) which contain one module, such as a gene promoter, a coding sequence, fusion tags and transcription terminators (Patron et al., 2015; Weber et al., 2011). The assembly of vectors by Golden-Gate relies on the joint activity of type-IIS restriction enzymes and T4 DNA ligase. In this way, in a single incubation L0 are digested with *Bsa*I, the products hybridise guided by the sticky-end-specificity which direct the order of assembly, and then ligated by T4 to form an L1. In a similar way, L1 are digested using *Bpi*I, they hybridise and ligate to form L2.

To perform a Golden-Gate reaction, 100 ng of purified L0 (or L1) were mixed with 100 ng of backbone donor, 1 µL of *Bsa*I (or *Bpi*I; 10 units/µL), 1 µL of T4 DNA ligase (5 units/µL; Thermo Scientific), 0.15 µL of BSA protein (20 mg/mL, NEB, only for L1 assembly), 1.5 µL of T4 DNA Ligase buffer with ATP 10X (Thermo Scientific) and topped to 15 µL with nuclease free H_2O (Sigma-Aldrich). The mix was incubated in a Veriti thermal cycler (Applied Biosystems) with the program shown in Table 2.7. The L1 (or L2) vector produced was then transformed in *E. coli* and selected using the corresponding antibiotics.

Table 2.7: Golden-Gate incubation program.

step	temperature (°C)	time (min)	cycles
1	37	3	25
	16	4	
2	50	5	1
	80	5	

2.3.2.4 *Agrobacterium*-mediated transformation of *Arabidopsis* (Floral dip)

A colony of *A. tumefaciens* containing the vector for transformation was used for inoculation of a 5 mL of LB medium supplemented with Kan, Rif and Gen for selection. The culture was grown overnight at 30 °C. 3 mL of this culture was used for inoculating a 50 mL LB medium with selection antibiotics. The culture was grown in a 250 mL conical flask shaking at 250 rpm for about 3 h at 30 °C until an OD_{600} between 0.6 and 0.8. The cells were sedimented at 3200 RCF for 5 min and resuspended in a 5% sucrose (w/v) solution. 0.005% Silwet L-77 (v/v) was added to the suspension. *Arabidopsis* plants were grown in soil for about 4 weeks, or until they developed a main stem with fully developed flowers. The flowers were fully submerged in the cell suspension for 2 min. Then the plants were covered with closed plastic transparent sleeves, and after three days the sleeves were opened. The plants were then dried, and their seeds were collected. Transformant seeds were selected by GFP fluorescence on the seed oil body membrane.

2.4 Cell wall xylan analysis

2.4.1 AIR preparation

The material was ground by ball milling in 96% ethanol. The milling was carried out in a Retsch MM400 mill in pulses of 10 min at 23 rpm separated by 5 min gaps. The pellet was washed with 99% ethanol, and then incubated with 2:3 (v/v) chloroform:methanol at room temperature for over 3 h. The supernatant was discarded and the pellet was incubated again with chloroform:methanol solution for over 3 h. Then the pellet was successively washed with 65%, 80% and 100% ethanol. The pellet was then dried at 45 °C for 2 days.

2.4.2 Holocellulose preparation

The amount of starting material varies between uses, but the ratios between material and buffers were maintained. Depectination was performed as follows: every 100 mg of AIR sample was suspended and incubated in 5 mL of 0.5% ammonium oxalate for 2 h at 85 ˚C with vigorous shaking. Then the sample was centrifuged for 10 min at 3200 RCF and washed with 10 mL of distilled H_2O. For young stem and callus materials, the holocellulose was ready after washing again with 10 mL of H_2O.

Mature stem material was additionally delignified. For this, the samples were incubated in 4 mL of 11% peracetic acid at 85 °C for 25 min. Then the samples were then chilled on ice and washed 4 to 5 times with H_2O until it reached a pH ~6. Then the supernatant was thoroughly removed with a pipette, then centrifuging the sample for 10 min at 3200 RCF and removing the extra supernatant accumulated.

2.4.3 Hemicellulose chemical extractions and treatments

2.4.3.1 Quick alkali extraction

Every 1 mg of AIR sample was immersed in 5 µL of 4 M NaOH and incubated for 1 h at 21 °C. The pH was adjusted to around 6 by adding ~20 µL of 1 M HCl and monitoring with litmus paper.

2.4.3.2 Overnight alkali extraction

5 g of AIR were immersed in 100 mL of a solution containing 6 M NaOH and 0.1% $NaBH_4$ (sodium borohydride). The suspension was incubated in a 1 L conical flask, overnight at 37 °C with vigorous shaking. Then the sample was chilled on ice and neutralised with 55 mL of glacial acetic acid. The polysaccharides were precipitated by adding 170 mL of 100 ethanol. The mixture was incubated on ice until flocculation. Then it was centrifuged at 3200 RFC for 10 min and the supernatant was discarded. The pellet was washed twice with 70% ethanol, snap-frozen in liquid nitrogen and lyophilised by sublimation.

2.4.3.3 Acetylated xylan extraction with DMSO

Freshly made holocellulose from 100 mg of AIR material was mixed with 5 mL of DMSO. The mixture was incubated in 50 mL tubes at 80 °C overnight with vigorous shaking. The supernatant was collected and evaporated under vacuum. When used for enzymatic digestions, the solid residue was then resuspended in 10 mL of 100 mM AmAc pH 6, ready for digestion.

When used for NMR, the solid residue was resuspended in 2.5 mL of 3 M urea, purified through a PD-10 column (General Electric, 17085101) and eluted with 3.5 mL of 42 mM urea. The sample was then snap-frozen and lyophilised by sublimation. The sample was then resuspended in 500 µL of deuterium oxide (D_2O), ready for NMR analysis.

2.4.3.4 Quick alkali treatment for extracted acetylated xylan

Xylan or extracted with DMSO or acetylated xylooligosaccharide samples (equivalent to up to 5 mg of AIR) were lyophilised by evaporation. The pellet was then covered with 5 µL of 4 M NaOH and incubated for 20 min at 21 °C. The pH was adjusted to around 6 by adding ~20 µL of 1 M HCl and monitoring with litmus paper.

2.4.4 Enzymatic tools

Enzymatic hydrolysis of xylan were carried out by using specific glycosyl hydrolases on acetylated xylan or xylooligosaccharides, either alkali-treated or non-acetylated.

Xylan digestions were carried out in single-pot reactions containing either substrate equivalent to 500 µg, 1 mg, 3 mg or 5 mg of AIR. The material was suspended in 250 µL of 100 mM AmAc pH 6 and pulverised with a micropestle. The enzymes used were added to the solution and incubated for the respective times at their specific temperatures (see Table 2.8). After the incubation, the enzymes were inactivated by incubating the mix at 100 °C for 10 min. GH11 and CE4 required an extra inactivation step by filtering the mixture through a Nanosep 3K Omega column (Pall, OD003C34).

The enzymes shown in Table 2.8 were obtained from different sources. BoGH30, GH11 and GH10 were kindly provided by Harry Gilbert (Newcastle, UK). EcGH30,

GH3, GH62 and GH115 were a gift from Novozymes. CE4 (CZ0034) and CtGH30 (CZ0445) were purchased from NZYTech.

Table 2.8: Parameters of carbohydrate-active enzymes used.
*the incubation times for BoGH30 and EcGH30 was of 30 min on glucuronoxylan and of 24 h on non-glucuronidated/acetylated samples.

Enzyme	Other names	Activity	Species	Concentration	amount per mg of AIR	Incubation
EcGH30	XynA	glucuronoxylanase	*Erwinia chrysanthemi*	1.6 mg/mL	0.3 µL	21 °C *
BoGH30	XynC2	glucuronoxylanase	*Bacteroides ovatus*	3 µM	10 µL	21 °C *
CtGH30	Xyn30A	glucuronoxylanase	*Clostridium thermocellum*	0,5 mg/mL	1 µL	30 min at 21 °C
GH11	GH11A	β-1,4-endoxylanase	*Neocallimastix patriciarium*	270 µM	5 µL	24 h at 40 °C
GH10	Xyn10A	β-1,4-endoxylanase	*Celvibrio japonicus*	5.2 mg/mL	10 µL	24 h at 21 °C
GH3		β-1,4-xylosidase	*Trichoderma reesei*	5.24 mg/mL	2 µL	3 h at 21 °C
GH115		α-1,2-glucuronidase	*Talaromyces sp.*	0.1 mg/mL	10 µL	24 h at 55 °C
GH62	α-1,2-arabinofuranosidase α-1,3-arabinofuranosidase		*Penicillum aurantiogriseum*	3.84 mg/mL	1 µL	24 h at 21 °C
CE4	Axe4A	Xylan acetylesterase	*Clostridium thermocellum*	1.5 mg/mL	2 µL	24 h at 40 °C

2.4.5 Reducing-end labelling (Reductive amination)

The RE of xylan and xylooligosaccharides possesses a reactive aldehyde group. The aldehyde group was chemically modified with four different chemical groups which are listed below.

2.4.5.1 ANTS

Derivatisation with ANTS (8-aminonaphthalene-1,3,6-trisulfonic acid, Invitrogen) was used for the visualisation and analysis of oligosaccharides by PACE. The procedure was carried out as described by Goubet et al. (2002), but 2-picoline-borane (2-PB) was used instead of cyanoborohydride. To derivatise oligosaccharides and standards (X_{1-6}, Megazyme) with ANTS, they were first lyophilised by evaporation. Then they were covered with 20 µL of ANTS derivatising solution (50% (v/v) DMSO, 7.5% (v/v) acetic acid, 10 mM ANTS, 10 mM 2-PB). The samples were briefly vortexed and incubated overnight at 37 ºC. Then the samples were lyophilised by evaporation at 55 ºC and resuspended in 3 M urea. The supernatant of the sample was then used for analysis on PACE.

2.4.5.2 APTS

Oligosaccharides were derivatised with 8-aminopyrene-1,3,6-trisulfonic acid (APTS, Sigma-Aldrich) for DASH analysis. The labelling procedure was implemented as described by Li et al. (2013). The reductive amination reaction was carried out by covering the sample with a mix containing 10 µL of APTS, 10 µL of 0.2 M 2-PB (in DMSO) and 10 µL of H_2O and incubating it overnight at 37 ºC. The resulting 30 µL of APTS labelled samples were then mixed with 970 µL of H_2O and considered the sample ‘Stock’ solution for the later DASH procedure.

2.4.5.3 2AA

Derivatisation with 2-aminobenzoic acid (2AA, Sigma) was used for the analysis of oligosaccharides by MALDI-MS. The derivatisation procedure was carried out using the optimised labelling conditions described by Ridlova et al. (2008). The oligosaccharides samples were first purified by using either a HyperSep column or a

C18 column (see below), and then dried. Then they were covered with 10 µL of a 2AA derivatisation mix and incubated for 3 h at 65 ºC. The mix contained 60 mg/mL of sodium cyanoborohydride (Sigma-Aldrich) and 60 mg/mL of 2AA in glacial acetic acid/dry DMSO (3:7; v/v). The labelled samples were subsequently purified from the derivatisation mix by using a GlycoClean S column (see below).

2.4.5.4 Procainamide

The labelling of oligossacharides with procainamide was also used for their MS analysis. The oligosaccharide samples purified with a HyperSep column were dried and dissolved in 100 µL of H_2O. The labelling was then performed as described by Lavanant et al. (2012). Procainamide hydrochloride (Sigma-Aldrich) was freshly dissolved to a concentration of 10 g/L in a solution containing 50% acetonitrile (AcN) and 1% orthophosphoric acid (procainamide solution). 1 µL of sample was mixed with 1 µL of procainamide solution and incubated at RT for 1 min. The mix was then directly used for mass-spectrometric analysis.

2.4.6 Oligosaccharides purifications

2.4.6.1 HyperSep Hypercarb column

For the analysis by MS, the digested xylooligosaccharides were first purified from the digestion buffer and enzymes by using a 1 mL HyperSep Hypercarb Porous Graphitized Carbon column (Thermo Scientific). The sample was first thoroughly lyophilised by evaporation to remove the AmAc buffer, and then it was resuspended in 1 mL of H_2O. The HyperSep column was connected to a peristaltic pump with a set flow of 0.5 µL/min. The column was purged by passing 2 mL of an 80% AcN solution acidified with 0.1% (v/v) trifluoroacetic acid. Then the column was equilibrated with 2 mL of H_2O. Then the sample was loaded into the column and washed with 1 mL of H_2O. Then it was washed with a 5% can solution. The sample was then eluted from the column by passing 400 µL of a 50% AcN solution acidified with 0.05% (v/v) trifluoroacetic acid. The eluted sample was then vacuum dried and ready for MS analysis.

2.4.6.2 Solid-phase extraction with C18 column

Sep-Pak Classic C18 cartridges (Waters, WAT051910) were used to remove the non-glucuronidated oligosaccharides from the samples. The C18 column was prepared by sequentially passing through it 15 mL of 100% ethanol, 20 mL of H_2O, 2 mL 100% AcN, and 10 mL of 20 mM AmAc pH 6. After the enzymatic hydrolysis, the sample was diluted to 1 mL with H_2O and loaded into the C18 column. Then the column was washed twice with 5 mL of H_2O (eluting the non-acetyalted oligossacharides). Then the sample was eluted with 5 mL of 10% ethanol (leaving the non-glucuronidated xylooligosaccharides in the column). The eluted fraction was snap-frozen in liquid nitrogen and lyophilised by sublimation.

2.4.6.3 GlycoClean S column

To remove the 2AA derivatisation mix from the labelled samples a GlycoClean S cartridge (ProZyme, GKI-4726) was used. The procedure was a modified version of the one described by Maslen et al. (2007). The column was sequentially washed with 1 mL of H_2O, 5 mL of 30% acetic acid, 4 mL of 100% AcN. Then the 10 mL sample was loaded onto the column and allowed to adsorb for 15 min. The column was sequentially washed with 1 mL of 100% AcN and 4 mL of 96% AcN. The purified sample was eluted from the column by passing 400 µL through the column three times. The samples were ready for MS analysis after being lyophilised by evaporation.

2.4.7 PACE

The digested oligosaccharides or entire xylan polymers were resolved by carbohydrate gel electrophoresis (PACE). The procedure used here is a modified version of the described by Goubet et al. (2002). Essentially, the 10% PACE gel composition had 19.2 mL of H_2O, 3 mL of 1 M Tris-Base Borate buffer pH 8.2 and 7.5 mL of 40% polyacrylamide (29:1 acrylamide/bis-acrylamide, Sigma, A7802). The 20% gel composition had 11.7 mL of H_2O, 3 mL of 1 M Tris-Base Borate buffer pH 8.2 and 15 mL of 40% polyacrylamide. Both gels were polymerised by adding to the mix 24 µL of TEMED (N,N,N',N'-tetramethylethane-1,2-diamine) and 240 µL of 10% (w/v) ammonium persulfate. No stacking gels were used. To resolve the oligosaccharides,

2-10 µL of ANTS-labelled sample were loaded in one of the wells of the gel. Additionally, 1 µL of a xylooligossacharide standards mix (5 mM of each X_{1-6}, Megazyme) was also loaded into another well of the gel. The gel was then exposed to a potential difference, electrophoresing the samples. 10% gels were electrophoresed at 1000 V for 43 min at 10 ºC. 20% gels were electrophoresed for 2.5 h instead. Labelled products on the PACE gel were detected by fluorescence using a G-Box CCD camera (Syngene) equipped with a transilluminator with long-wave tubes emitting at 365 nm. Images were captured using GeneSnap software (Syngene). Densitometric quantifications were carried out using ImageJ (National Institutes of Health) and SnapTools (Syngene) softwares.

2.4.8 DASH

DASH was used for the quantification of the percentage of methylation of the GlcA substitutions of xylan (Li et al., 2013). APTS-derivatised 'Stock' samples were subjected to a series of sequential 10-fold dilutions. Different amounts, from 5 µL to 120 µL of 1/1000 to 1/100000 dilutions were loaded onto a 96-well plate. 2 µL of fractional mobility marker stock solutions were added to each well for a later alignment of the resulting spectra (as described). The samples on the plate were lyophilised by evaporation. 20 µL of DMSO was added to each sample. Then the plate was vigorously shaken for 1 h at RT. The plates were analysed on a 96-capillary 3130XL DNA analyser (Applied Biosystems) by the DNA Sequencing Facility of the Department of Biochemistry (University of Cambridge). DASH data was then analysed using DASHBOARD (Li et al., 2013).

2.4.9 Mass Spectrometry (MS)

Purified unlabelled, 2AA-derivatised or procainamide-derivatised samples were submitted to Oligosaccharide mass profiling (OLIMP) analysis by mass-spectrometry. The procedure was carried out following the protocol presented by Tryfona et al. (2012) with some modifications.

2.4.9.1 MALDI-ToF/MS

Desalted or 2AA labelled oligosaccharides were resuspended in 100 µL of H_2O, and mixed in a 1:1 (v/v) ratio with matrix. The matrix solution had 10 mg/mL 2,5-dihydroxybenzoic acid (DHB, Sigma-Aldrich) and 0.4 mg/mL ammonium sulphate ($(NH_4)_2SO_4$); to prevent formation of di-sodiated adducts) in 50% aqueous methanol (Enebro and Karlsson, 2006). For procainamide-labelled samples, the sample solution was mixed in a 1:1 ratio with matrix. 2 µL of sample/matrix mix were spotted on a MALDI MTP 384 ground steel BC target plate (Bruker).

Sample spots were rapidly dried in a vaccum desiccator. All mass-spectra were recorded on an UltrafleXtreme matrix-assisted laser desorption ionization-time of flight/time of flight (MALDI/TOF-TOF) instrument (Bruker). The spectra were collected with a 2-kHz smartbeam-II laser on reflector mode. The mass range was 300-3000 Da, with a signal suppression of 150 Da. Averages between 10000 and 50000 shots were used to obtain the MS spectra. FlexControl and FlexAnalysis softwares were used for data acquisition and analysis (Bruker).

2.4.9.2 HILIC and MALDI TOF/TOF-MS/MS

Normal-phase hydrophilic-interaction liquid chromatography (HILIC) was carried out using acetylated xylan digestion products. Mature and young stem digestion products were first purified by solid-phase extraction with a C18 column, and then labelled with 2AA. For callus digestion products, the samples were desalted with a HyperSep column and then labelled with 2AA. The samples were then purified with a GlycoClean S Column and lyophilised. One sample was then resuspended in 20 µL of H_2O and mixed with 4 volumes of 100% AcN (Rathburn, RH1016). Capillary HILIC was carried out using an LC-Packings Ultimate system (Dionex, www.dionex.com) equipped with an amide-80 column (300 µm x 25 cm, 5 µm bead size; LC Packings/Dionex) and with a capillary HPLC system. The flow was set at 3 µL/min: Solvent A was 50 mM ammonium formate adjusted to pH 4.4 with formic acid. Solvent B1 was aqueous 80% AcN with 10 mM ammonium formate adjusted to pH 4.4 with formic acid. Solvent B2 was aqueous 95% AcN with 2.5 mM ammonium formate adjusted to pH 4.4 with formic acid. Solvent B1 was used for HILIC separation of callus samples, and solvent B2 for mature and young stems samples.

2AA-labelled samples dissolved in 80% AcN were loaded onto the column and eluted with increasing aqueous concentration. The column effluent was directly connected to a probot sample fractionation system and spotted onto an MTP SmallAnchor 384 BC target plate (Bruker) at 20 s intervals. The spotted fractions were air-dried, and then, each spotted fraction was overlaid with 0.5 µL of DHB matrix and air-dried. The data acquisition was performed in the same MALDI mass-spectrometer described in section 2.4.9.1.

The oligosaccharide molecular ions were identified in the MALDI data and their elution positions were determined with the WARP-LC software (Bruker). Molecular ions were selected for LASER-induced dissociation (LID) ToF/ToF MS/MS analysis using the LIFT apparatus (for birch spectrum see next paragraph). LIFT spectra were acquired by averaging 2000 shots of parent ion spectra, averaging between 10000 and 50000 shots of fragment spectra, and summing both averages using the FlexControl software (Bruker).

The birch phloem xylan oligosaccharide sample was exceptionally analysed by directly spotting the sample onto the MALDI plate after 2AA derivatisation and purification by Glycoclean S cleanup. The spectrum of only this sample was acquired by collision-induced dissociation (CID) ToF/ToF MS/MS. For this the CID cell of the MALDI apparatus was filled with argon at a pressure of 7×10^{-6}, prior to the acquisition of the spectrum.

2.4.10 Nuclear magnetic resonance (NMR)

The NMR experiments of WT, *tbl32*, *tbl33* and *tbl32/33* were carried out in collaboration with T. Almeida and Prof K. Stott. Acetylated xylan was extracted from 200 mg AIR with DMSO after depectination and delignification pretreatments. After the xylan purification with PD-10, the sample was resuspended in 0.5 µL of D_2O with a final concentration of 0.5 M urea. A Bruker ADVANCE III spectrometer operating at 600 MHz equipped with a TCI Cryo-probe was used to obtain NMR spectra by T. Almeida. 1D-NOESY (Nuclear Overhausser Effect Spectroscopy), $^{1}H/^{13}C$-HSQC (Heteronuclear single quantum correlation coherence), 2D-$^{1}H/^{1}H$-NOESY and 2D-$^{1}H/^{1}H$-TOCSY (Total correlation spectroscopy) experiments were all carried out at 35 °C. The data processing was performed by T. Almeida using TopSpin version 3.7 (Bruker) and AZARA 2.8 (copyright 1993-2014, Wayne Boucher and Department of Biochemistry, University of Cambridge) were used. Peaks were manually assigned by T. Almeida using CCPN v2.42, by peak correlations between experiments, or following chemical shifts previously reported (Bromley et al., 2013; Grantham, 2016). *gxm1/2/3* acetylated xylan was used for the assignment of some non-methylated peak correlations, although this is not shown.

Chapter 3: Acetylation and GlcA substitution patterns of secondary-cell-wall xylan in xylan backbone synthesis mutants

3.1 Introduction

In *Arabidopsis* (and other eudicots) the secondary-cell-wall xylan is synthesised in the Golgi apparatus by a number of diverse enzymes (Hao and Mohnen, 2014). The backbone is first elongated, and then it is substituted with Ac and GlcA by acetyltransferases and glucuronyltransferases, respectively. However, the order of these processes during the synthesis, and the way they are coordinated has not been fully determined.

Three types of protein in the XSC are each necessary for xylan elongation; IRX9/9-L, IRX10/10-L and IRX14/14-L (see Chapter 1, section 1.3.3, page 16) (Brown et al., 2009; Wu et al., 2010). For xylan synthesis in primary cell walls, IRX9-L has been shown to be necessary and therefore could be considered a primary wall XSC component, but in secondary-cell-wall xylan synthesis, IRX9 and its homologue IRX9-L are both capable of contributing to xylan synthesis (Lee et al., 2010; Wu et al., 2010). However, in the *irx9* mutant where only IRX9-L is active, there is less xylan in the secondary wall, and also the length of the xylan chains is shorter (Peña et al., 2007). In the absence of both IRX9 and IRX9-L, there is no xylan synthesis, causing the plant growth to be severely dwarfed (Wu et al., 2010). A similar situation occurs for the pairs IRX10/IRX10-L and IRX14/IRX14-L. When IRX10 (or IRX14) is knocked out, IRX10-L (or IRX14-L) can still direct xylan synthesis, but the xylan chains are shorter (Brown et al., 2009, 2007). IRX10-L is responsible alone for primary-wall xylan synthesis, and could be considered also a primary wall XSC component (Urbanowicz et al., 2014). IRX14 and IRX14-L seem to be slightly less specialised: both are capable of producing primary and secondary-cell-wall xylan in the absence of the other (Brown et al., 2007; Mortimer et al., 2015; Wu et al., 2010). Two putative methyltransferases, IRX15 and IRX15-L, also affect xylan backbone elongation (Brown et al., 2011; Jensen et al., 2011). In *irx15/15-l* the xylan chains are shorter in length, but unlike the other *irx* mutants, the xylosyltransferase activity of the XSC in this mutant is not substantially decreased, which suggests they are not an essential part of the XSC.

IRX10 and IRX10-L are thought to possess the active site in the XSC, because they can transfer Xyl to xylooligosaccharides *in vitro* (Jensen et al., 2014; Urbanowicz et al., 2014). However, the functions of IRX9, IRX9-L, IRX14 and IRX14-L in the xylan synthesis are not fully understood.

Preliminary results (J. Bromley, P. Dupree, unpublished results) suggested that in the xylan of *irx9*, *irx10* and *irx14*, the pattern of [Me]GlcA might be disrupted, indicating that the xylan produced by IRX9-L, IRX10-L and IRX14-L may have altered glucuronidation. This suggests that perhaps GUX1 or GUX2 might be dysregulated. If so, there might be a relationship between the process of xylan backbone elongation and the xylan glucuronidation, and this link may be affected when the composition of the XSC is altered. Moreover, the pattern of glucuronidation in primary-wall xylan, synthesised by IRX9-L, IRX10-L, IRX14/14-L and GUX3 (Mortimer et al., 2015), is different (see Chapter 1, Figure 1.22, page 34). This raises the questions whether the XSC with a primary-cell-wall-xylan-synthesis subunit produces a different [Me]GlcA pattern of xylan.

Since GUX1 and GUX2 produce a different pattern of xylan (Bromley et al., 2013), it is possible that only one of the patterns may be affected by the inactivation of IRX9, IRX10 and IRX14. GUX1 and GUX2 perform the glucuronidation of the xylan in secondary cell walls (Mortimer et al., 2010). In *Arabidopsis*, GUX1 decorates about 70% of the xylan, and the GlcA that GUX1 transfers to the xylan backbone are spaced by an even number of Xyl (Bromley et al., 2013). The predominant distances between GUX1-mediated-[Me]GlcA substitutions are 8, 10 and 12 Xyl. In contrast, GUX2-mediated-[Me]GlcA are spaced by either even or odd number of backbone residues, and are spread on about 30% of the xylan. The distances between GUX2-mediated-[Me]GlcA are most frequently spaced by 5, 6 and 7 backbone Xyl. If the activity of GUX1 or GUX2 is linked to backbone synthesis, then their activity might be affected by alteration of the XSC composition. If so, the [Me]GlcA pattern produced by either, or both GUX enzymes may be affected.

Perhaps in *irx9*, *irx10* and *irx14*, the region of the xylan backbone that is decorated by GUX1 is shortened. If GUX1 is dysregulated, then it would be expected to find an alteration of the frequencies of spacing between [Me]GlcA on xylan in the *irx* mutants.

The inactivation of IRX9, IRX10 or IRX14 could have an alteration in the main acetylation pattern (Figure 1.21), which could additionally affect the downstream process of xylan glucuronidation. The Ac also follow a pattern of the secondary-cell-wall xylan backbone. Almost every other Xyl of the backbone is acetylated, and when [Me]GlcA on xylan is genetically removed (in *gux1/2*), the main acetylation pattern of xylan is essentially unaltered (Busse-Wicher et al., 2014; Chong et al., 2014). How this pattern is generated is unknown, but it might arise by coupling acetylation to the xylan backbone emerging from the XSC. ESK1 is responsible for transferring about 50% of the Ac to the xylan of secondary cell walls (Xiong et al., 2013). When the experiments in this chapter were planned and carried out, there was no published knowledge about the acetylation of primary-cell-wall xylan. Therefore, it was a possibility that such primary-wall xylan, synthesised by different XSC components, is acetylated by different enzymes, or acetylated with a different pattern or non-acetylated. If the main acetylation pattern is modified in the mutants, this could suggest a mechanistic link between the XSC and xylan acetylation.

An apparent effect on the xylan [Me]GlcA pattern in the *irx* mutants could also be caused by the short length of the xylan chain. To determine the [Me]GlcA pattern of xylan, the reported method uses the activity of GH30 glucuronoxylanase (Bromley et al., 2013). This enzyme hydrolyses the xylan once per [Me]GlcA substitution, and the number of backbone Xyl on the digestion products is the distance between two consecutive GlcA. Therefore, the relative abundance of the GH30 products on xylan reveals the frequency of the distance between [Me]GlcA. Additionally, at the ends of the molecule, the oligosaccharide arises from the [Me]GlcA nearest the end. In WT the xylan backbone is about 120 Xyl long (Busse-Wicher, unpublished data), and a mean distance between [Me]GlcA substitutions of 8 Xyl approximately (Bromley et al., 2013). Then, for each xylan molecule there are on average 15 [Me]GlcA substitutions. In mutants in which the xylan is shorter, some chains are 25 Xyl long (Busse-Wicher, unpublished data), then each chain would have on average 3 [Me]GlcA modifications. The interpretation of the [Me]GlcA pattern of WT and the mutants' xylan would be determined by 25 and 3 fragments produced by GH30 respectively, but in both cases, the non-specific oligosaccharides originated from the ends of the molecule would be similar. Therefore, in the short-xylan case the signal/noise ratio is reduced to about

12% respect to normal-length xylan, which may impact on the resolution for the detection of the pattern.

The aim of this chapter was to investigate whether the altered composition of the secondary cell wall XSC in *irx9*, *irx10* and *irx14* has an effect on the substitution patterns of xylan. To analyse the [Me]GlcA pattern in these mutants, de-acetylated xylan was digested with a GH30 glucuronoxylanase (Bromley et al., 2013). To analyse the effects on xylan acetylation caused by the inactivation of IRX9, IRX10 or IRX14, the [Me]GlcA modifications were genetically removed by crossing with *gux1/2*. DMSO-extracted acetylated xylan was digested with GH10 which reveals patterns of acetylation (Busse-Wicher et al., 2014). Finally, to test further if secondary-cell-wall xylan substitution enzymes are affected by the identity of the proteins in the XSC or the quantity of activity, IRX9-L was expressed under the IRX9 promoter in secondary cell walls and the substitution patterns were analysed.

3.2 Results

3.2.1 The xylan [Me]GlcA pattern is altered in *irx9*, *irx10* and *irx14*

To determine the xylan GlcA pattern in the XSC mutants, their xylan was analysed using the glucuronoxylanase GH30 as described (Bromley et al., 2013). To include the short-xylan effect in the analysis, *irx15/15-l* xylan was used as a short-xylan control, with a normal XSC composition. AIR was prepared from WT, *irx9*, *irx10*, *irx14* and *irx15/15-l* basal stems and the xylan was extracted with alkali. The xylan was then digested with GH30, and the products were labelled with the fluorophore ANTS and resolved by PACE (Figure 3.1A). The bands that correspond to $U^{[me]}X_{5\text{-}10}$ were quantified from the gel to aid interpretation (Figure 3.1B).

In WT xylan, GH30 produced xylooligosaccharides mostly with even number of backbone residues, and the most abundant was $U^{[me]}X_8$. This is characteristic of the GlcA pattern produced by GUX1. In contrast, $U^{[me]}X_5$, $U^{[me]}X_7$ and $U^{[me]}X_9$ were less abundant than $U^{[me]}X_6$, $U^{[me]}X_8$ and $U^{[me]}X_{10}$ respectively.

The XSC mutants' xylan exhibited less preference for the even length GlcA pattern than xylan from WT. *irx10* xylan digestion mostly produced oligosaccharides with an

even number of backbone residues, but the products with odd number of backbone residues were more represented. The relative abundance of $U^{[me]}X_5$ and $U^{[me]}X_7$ was much closer to the predominant products. From *irx9* xylan, the most abundant GH30 product was $U^{[me]}X_6$, which has even number of backbone Xyl, but the predominance between $U^{[me]}X_5$, $U^{[me]}X_7$ and $U^{[me]}X_8$ was not clear. The difference in abundance between $U^{[me]}X_9$ and $U^{[me]}X_{10}$ was greatly reduced in *irx9* compared to WT. Interestingly, the predominant GH30 product on *irx14* was $U^{[me]}X_5$, which has an odd number of backbone residues. The abundance of $U^{[me]}X_6$ from *irx14* predominated over $U^{[me]}X_7$, which dominated over $U^{[me]}X_8$, and this over $U^{[me]}X_9$. The order in predominance of GH30 products, and the predominance of a species with odd number of backbone Xyl suggests that the [Me]GlcA pattern of xylan from WT would be most distinct from *irx14* from all the mutants analysed. The higher abundance of $U^{[me]}X_8$ and $U^{[me]}X_{10}$ in *irx10* suggests that the [Me]GlcA pattern in *irx10* xylan would be less altered than *irx9*. It would be useful to perform additional replicates for this analysis for an accurate quantification.

In *irx15/15-l*, oligosaccharides with even number of backbone Xyl were also predominant, and species with odd number of backbone Xyl were also produced. However, $U^{[me]}X_8$ was not the most abundant GH30 product, possibly because in *irx15/15-l* the xylan is shorter. *irx15/15-l* exhibited a lower alteration of the xylan [Me]GlcA pattern than the other mutants. If the size of the xylan chain is similar between all mutants, this suggests that the artefact produced by a short xylan chain does not explain the alteration on the [Me]GlcA pattern in *irx9*, *irx10* and *irx14*. It is also possible that in the XSC mutants the xylan may be shorter than in *irx15/15-l*.

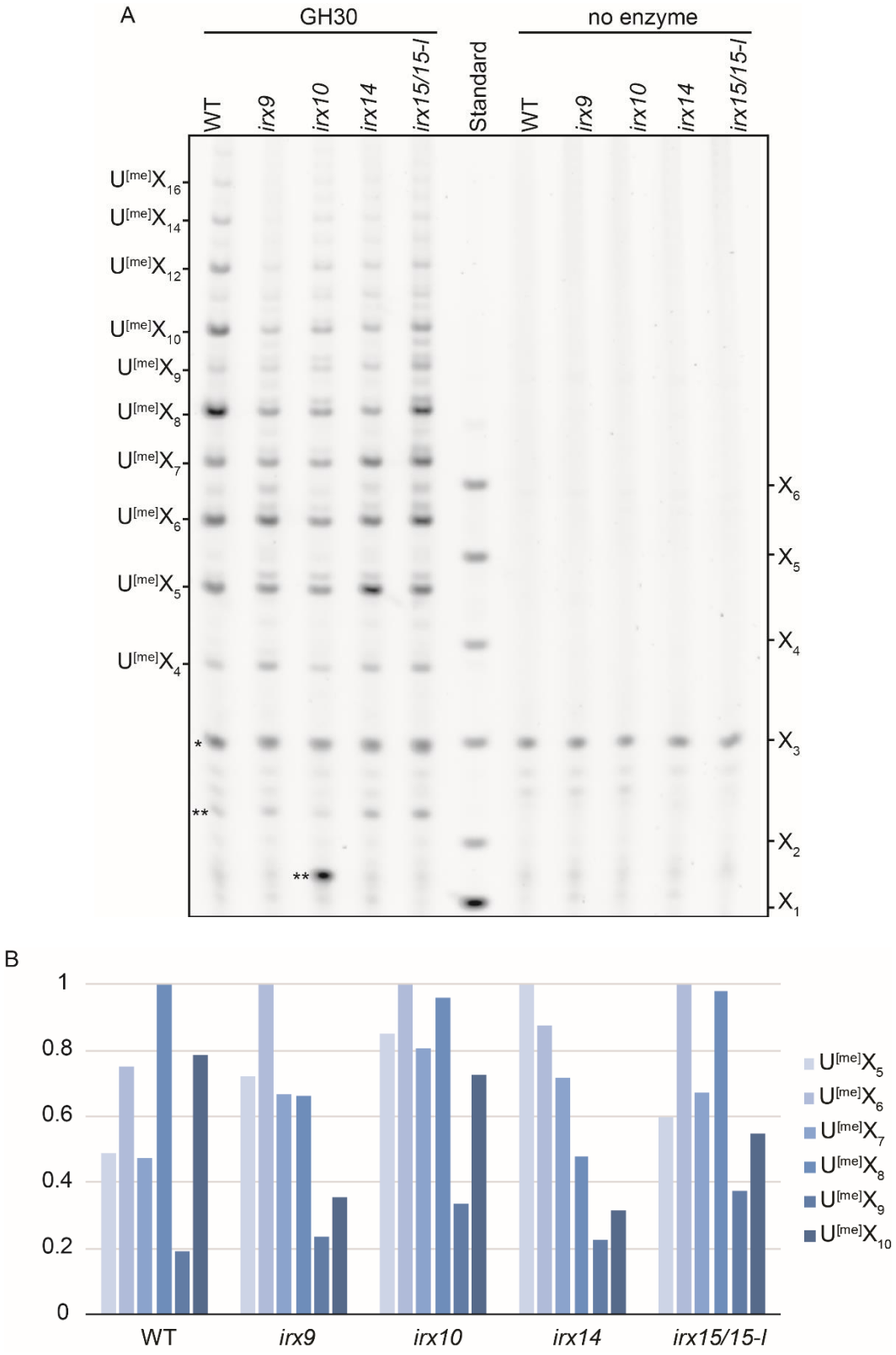


Figure 3.1: GH30 digestion of the *irx* mutants.
Xylan was extracted with alkali from AIR and digested with the glucuronoxylanase BoGH30. (A) The digestion products were labelled with ANTS and resolved by PACE on a 20% polyacrylamide gel. *contaminant, ** unidentified. (B) The intensity of the bands on this gel corresponding to $U^{[me]}X_{5\text{-}10}$ were quantified, normalised to the maximum within the sample. One biological replicate.

3.2.2 The inactivation of IRX9, IRX10 or IRX14 does not alter the acetylation pattern of xylan

It is likely that the xylan acetylation occurs shortly after the backbone is elongated. If the elongation of the xylan chain and the acetylation process are mechanistically linked, altering the composition of the XSC by inactivating IRX9, IRX10 or IRX14, may cause alterations in the acetylation pattern of xylan. The effects on the acetylation were studied by analysing the products of GH10 xylanase on acetylated xylan. GH10 products on WT acetylated xylan have [Me]GlcA in multiple positions, which is complex to interpret, but when [Me]GlcA are absent, GH10 produces a ladder of oligosaccharides, and the length and abundance of these oligosaccharides provides information about the pattern of spacing between Ac (Busse-Wicher et al., 2014). Therefore, it was convenient to genetically remove the [Me]GlcA in the mutants.

The xylan [Me]GlcA substitutions were genetically removed by producing the mutants *irx9 gux1/2*, *irx10 gux1/2*, *irx14 gux1/2* and *irx15/15-l gux1/2* (*irx gux1/2* mutants). The crosses between *irx9*, *irx10*, *irx15/15-l* and *gux1/2* were performed by X Yu and M Busse-Wicher. The selection of homozygous triple mutants was performed as part of the work in this thesis by genotyping using PCR as described in Chapter 2 (sections 2.3.1.2 and 2.3.1.5). The lack of GlcA in the mutants was verified by PACE using GH11 and CE4 (see annex, Figure S1). The absence of $XU^{[me]}XX$ in the GH11 products was diagnostic of the absence of GlcA in the mutants' xylan. Nevertheless, *gux1/2* and the generated mutants possess [Me]GlcA in their xylan from primary cell walls. To identify GH10 products containing [Me]GlcA that may originate from primary cell walls, the triple mutant *gux1/2/3*, was also included as a control in the analysis (Mortimer et al., 2015, 2010). There was no obvious change in growth phenotype in the *irx gux1/2* mutants, in comparison to the reported phenotypes of the *irx* mutants (Figure 3.2) (Brown et al., 2011, 2009, 2007; Wu et al., 2010). This suggests that the addition of mutations in *GUX1* and *GUX2* to the *irx* mutants does not produce less functional xylan.

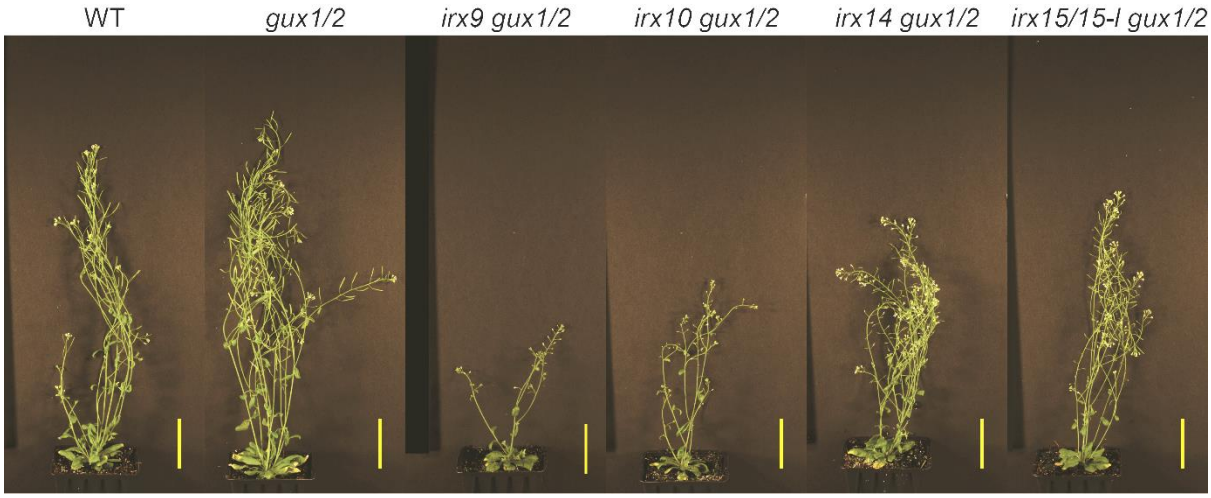


Figure 3.2 Growth phenotypes of *irx gux1/2* mutants.
Plants were grown for 6 weeks. Representative plants are shown for each genotype. Scale bar = 5 cm.

As an additional control, to verify that the IRX were inactivated, the length of xylan was studied in the *irx gux1/2* mutants. For this, acetylated xylan was extracted with DMSO, derivatised with ANTS and analysed by PACE (see Figure 3.3). The xylan relative size is reflected by the distribution of the smear produced on the gel. However, the charge contributed by [Me]GlcA can increase the mobility of the xylan molecules. Because there is no effect on the mobility of *gux1/2* and *gux1/2/3* xylan on PACE caused by the [Me]GlcA negative charge, they are more suitable controls than WT xylan for comparing the polymer length in the *irx gux1/2* mutants. The xylan of *gux1/2* and *gux1/2/3* is depleted of [Me]GlcA substitutions, but the chain lengths are similar to WT (Grantham, 2016; Mortimer et al., 2010). Interestingly, the xylan of all *irx gux1/2* mutants migrated much further into the gel than the controls, indicating that they have a reduction in xylan chain length. *irx9 gux1/2* xylan migrated faster than the xylan of the *irx10 gux1/2* and *irx14 gux1/2* mutants. This is consistent with the view that *irx9* has shorter xylan than *irx10* and *irx14* (Busse-Wicher, unpublished data). This result confirmed that the *irx gux1/2* mutants lack the normal activity of the respective IRX protein.

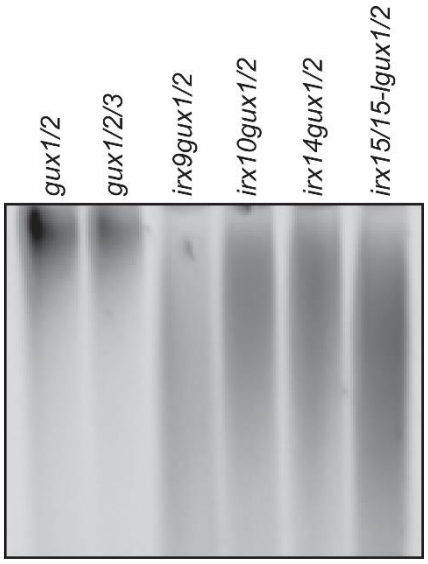


Figure 3.3 Xylan length analysis by PACE of the *irx gux1/2* mutants.
Acetylated xylan was extracted with DMSO and derivatised with ANTS. The undigested xylan was separated by gel electrophoresis in a 10% polyacrylamide gel.

The xylan acetylation pattern was next analysed in the *irx gux1/2* mutants. For this, AIR was prepared from basal stems and the xylan was extracted with DMSO. The extracted xylan was partially digested with GH10 xylanase. GH10 requires an unsubstituted Xyl on the -1 subsite, and can tolerate a modified Xyl on +1 subsite (see section 1.5.2). These GH10 specificities result in its products allowing a substituted NRE Xyl and always having an unsubstituted Xyl on the RE. Therefore, GH10 products on non-glucuronidated acetyl-xylan indicate the distance between two acetylated Xyl. After the hydrolysis with GH10, to improve the migration on the polyacrylamide gel, the Ac were removed from the oligosaccharides by treating them with alkali, then they were labelled with ANTS and resolved on a PACE gel (Figure 3.4A). The intensities of X_1, X_2, X_3, X_4, X_5 and X_6 for each sample on the gel were quantified (Figure 3.4B).

The xylan acetylation pattern in the *irx gux1/2* mutants was revealed by the relative abundance of the GH10 products. In the control *gux1/2*, X_2 was more abundant than X_1 and X_3; X_4 was more abundant than X_3 and X_5; and so on. This preference for the even number of Xyl of the digestion products implies that *gux1/2* xylan is acetylated mostly every other Xyl. This is consistent with the acetylation pattern in *gux1/2* xylan previously reported (Busse-Wicher et al., 2014). The GH10 products on *gux1/2/3* xylan co-migrated with the ones from *gux1/2*, indicating that there was no (methyl)glucuronidated xylan contamination from primary cell wall.

The acetylated xylan digestion pattern seen in xylan from *gux1/2* is remarkably reproduced by the xylan of *irx9 gux1/2*, *irx10 gux1/2*, *irx14 gux1/2* and *irx15/15-l gux1/2*. In all these mutants, the most abundant species produced by the hydrolysis with GH10 (and alkali treatment) was X_2, and the second most abundant species was X_4. Both of these species are produced from substrate xylan that has X^a residues separated by even number of backbone residues. Therefore, this implies that the main alternating acetylation pattern of xylan was unaffected by the inactivation of IRX9, IRX10 or IRX14. Additionally, the inactivation of IRX15/15-L did not cause an apparent alteration of the main acetylation pattern of xylan neither.

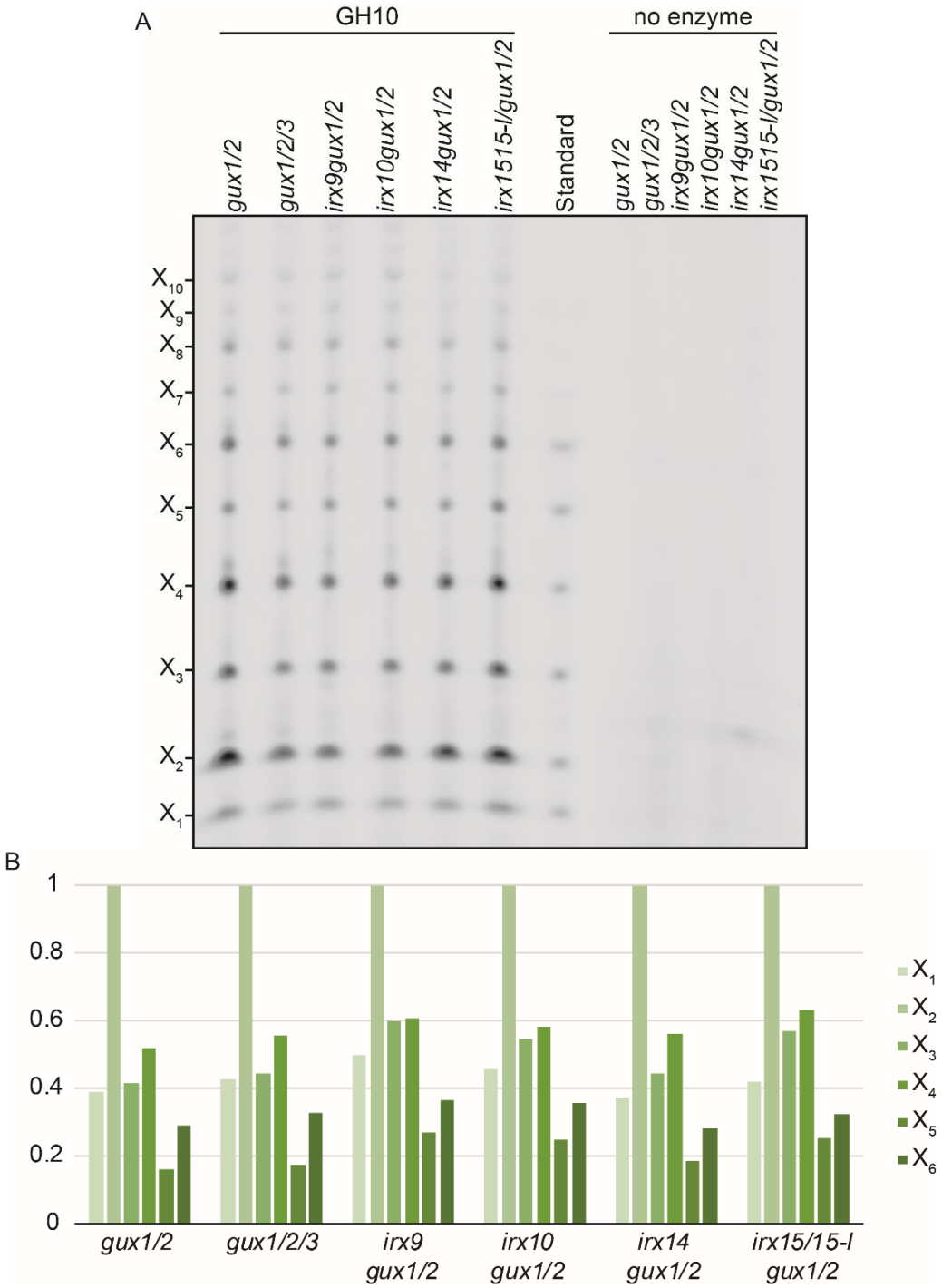


Figure 3.4 GH10 digestion of xylan from the *irx gux1/2* mutants.
Xylan was extracted with DMSO from AIR and digested with the xylanase GH10. (A) The digestion products were de-acetylated by treating them with alkali, then they were labelled with ANTS and resolved by PACE on a 20% polyacrylamide gel. (B) The intensity of the bands on this gel corresponding to $X_{1\text{-}6}$ were quantified, normalised to X_2, and graphed. Representative of two biological replicates

3.2.3 IRX9-L can produce full-length and patterned xylan

In *irx9*, the XSC with IRX9-L produces less xylan, with shorter chains (Wu et al., 2010). The overexpression of IRX9-L under the general 35S promoter in *irx9* restores WT growth size however. It is unknown whether growth was restored because a higher amount of short, defective xylan is produced. Or, is it because the xylan chain length and/or the xylan GlcA substitution pattern are recovered, restoring functionality?

To identify if IRX9-L can allow synthesis of full length xylan with a WT GlcA pattern, plants overexpressing IRX9-L in secondary cell wall synthesising cells were generated. For this, *irx9* mutants were transformed with a vector for the expression of IRX9-L under the secondary cell wall-specific promoter of IRX9 (Figure 3.5). The transformed plants were selected by observing green fluorescence on their seed oil body membrane produced by the Oleosin-GFP selection marker (Shimada et al., 2010). Homozygous transformed lines were isolated. The insertion of one copy of the transgene was verified by the segregation proportion 3:1 of fluorescent:non-fluorescent seeds in the offspring of hemizygous plants. Homozygous plants were verified by observing 100% fluorescent seeds of progeny in the following generation. Three representative lines were selected for analysis based on the apparent restoration of growth to WT size (Figure 3.6; L5, L6 and L8). It would be useful to verify expression of IRX9-L in these lines via Western Blot or RT-PCR analysis.

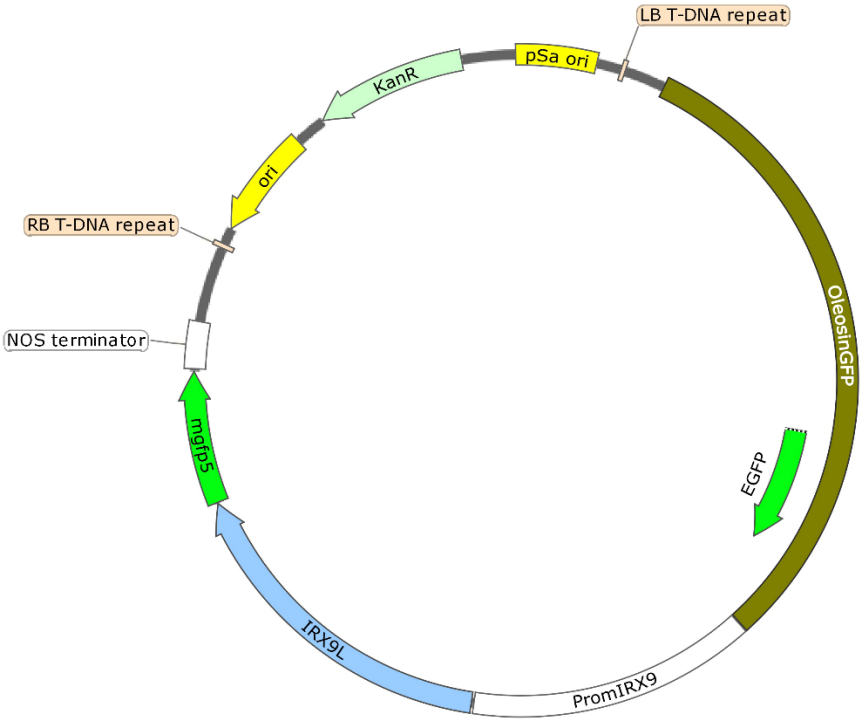


Figure 3.5 Vector for the overexpression of *IRX9-L* in *irx9* lines.
Genomic *IRX9-L* coding region was cloned between IRX9 promoter and *GFP* (*mgfp5*). The vector was assembled by conventional restriction enzyme cloning as described in Chapter 2 (section 2.3.2.2). OleosinGFP is a selection marker that expresses GFP in the seed. Size ~10.9 Kb.

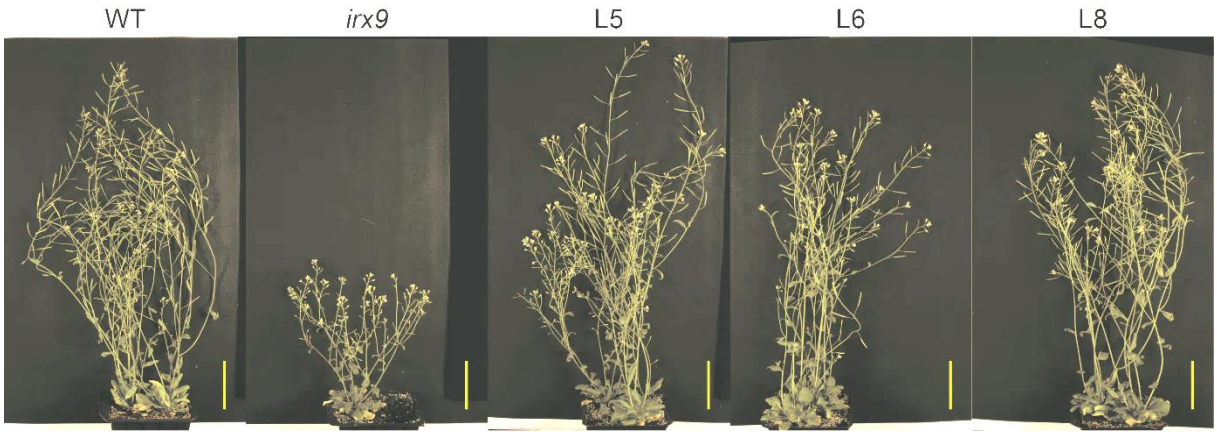


Figure 3.6: *irx9* lines overexpressing IRX9-L.
Three representative lines produced from independent transformations with the vector shown in Figure 3.5 for the overexpression of IRX9-L in secondary-cell-wall-producing tissues. Scale bar = 5 cm.

To analyse the xylan chain length, material was harvested from basal stems of the transformed lines and AIR was prepared. Xylan extracted from WT and untransformed *irx9* was used for normal and short xylan chain length controls, respectively. The xylan was then extracted with alkali, derivatised with ANTS and analysed by PACE (see Figure 3.7). The xylan length is qualitatively visible by the overall migration of the polymers from the top to the bottom of the gel. The WT full-length xylan produced a smear that migrated a short distance into the gel, while the shorter xylan of *irx9* migrated further down the gel. The xylan of L5, L6 and L8 had a mobility in the gel similar to the xylan of WT, and very different to that of the *irx9* mutant. The recovery of the wild-type xylan size phenotype in the transformed lines indicates that XSCs containing IRX9-L are capable of producing xylan chains of WT length.

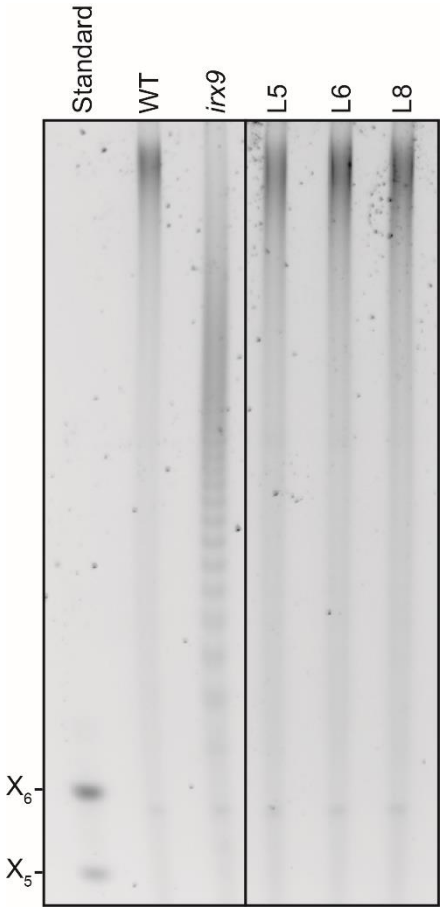


Figure 3.7 Xylan length analysis by PACE of IRX9-L overexpressors.
Xylan was extracted with alkali from *irx9* lines overexpressing IRX9-L under the IRX9 promoter. The intact xylan was labelled with ANTS and analysed by PACE. 10% polyacrylamide gel. Samples and controls were resolved in one gel, middle lanes were removed. One biological replicate for each transformed line.

Next, the [Me]GlcA substitution pattern of xylan was analysed in the overexpressor lines. For this, the alkali-extracted xylan from the transformed lines was enzymatically hydrolysed with GH30. WT and *irx9* xylans were used as normal and altered GlcA-pattern controls, respectively. Then, the xylan was labelled with ANTS and analysed by PACE (see Figure 3.8A). The band intensities for $U^{[me]}X_{5\text{-}10}$ on the PACE gel were measured, the values obtained were normalised to the highest within the sample, and graphed (Figure 3.8B).

The most abundant species produced by GH30 on WT was $U^{[me]}X_8$ followed by $U^{[me]}X_6$, both with even number of backbone residues. *irx9* produced predominantly $U^{[me]}X_6$ followed by $U^{[me]}X_5$, and $U^{[me]}X_8$ had a similar abundance to $U^{[me]}X_7$. The intensities of the bands produced by WT and *irx9* xylan were consistent with the [Me]GlcA distributions patterns described in section 3.2.1. Interestingly, the three lines *irx9* overexpressing IRX9-L (L5, L6 and L8) predominantly produced $U^{[me]}X_8$ followed by $U^{[me]}X_6$. Compared to WT, the difference in relative abundance between of $U^{[me]}X_8$ and $U^{[me]}X_6$ was reduced in all the overexpressor lines. In addition, the production of $U^{[me]}X_5$ seemed reduced in L5, L6 and L8 compared to WT and *irx9*. The transformed lines therefore rescued the *irx9* xylan GlcA patterning phenotype.

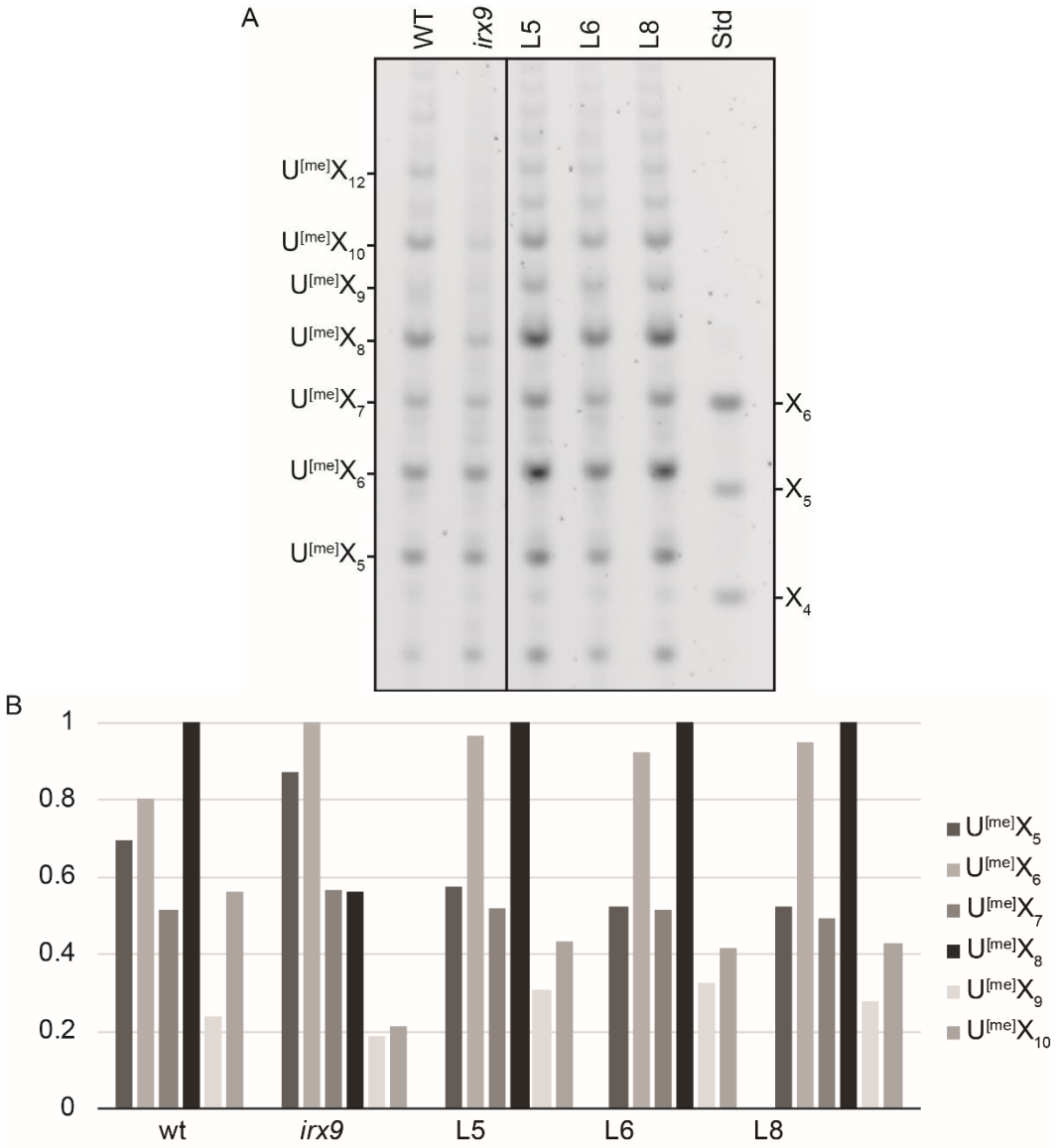


Figure 3.8 GH30 digestion of IRX9-L overexpressor lines in the *irx9* mutant.
AIR was prepared from basal stems, the xylan was extracted with alkali and digested with the glucuronoxylanase BoGH30. (A) The digestion products were labelled with ANTS and resolved by PACE on a 10% polyacrylamide gel. Samples and controls were resolved in one gel, middle lanes were removed. (B) The intensity of the bands on this gel corresponding to $U^{[me]}X_{5\text{-}10}$ were quantified, and are shown normalised to the maximum within the sample. One biological replicate for each transformed line.

3.3 Discussion

The main objective of this chapter was to investigate whether there might be a mechanistic interaction between the xylan backbone synthesis enzymes and the xylan acetylation and glucuronidation processes, and whether this interaction influences the patterning of the substitution on xylan. To test this idea, the xylan substitution patterns of secondary-cell-wall xylan from the *irx9*, *irx10* and *irx14* mutants were analysed. These mutants offer a unique opportunity to investigate if any IRX enzyme of the secondary cell wall XSC has important functions in determining xylan substitution patterns. In these mutants the secondary cell wall XSC member is replaced by a homologue that is required for primary-cell-wall xylan synthesis. The inactivation of IRX9, IRX10 and IRX14 caused an alteration of the GlcA pattern of xylan which was visible with PACE, but there was no evidence that the main alternating pattern of acetylation is affected. Overexpression of IRX9-L in secondary wall exhibited a partial recovery of the xylan [Me]GlcA pattern caused by the inactivation of IRX9. There was evidence that the patterning of xylan is influenced by the activity of the secondary-cell-wall-specific XSC subunits, and also that the patterning can be achieved using the primary-cell-wall xylan synthesis subunits IRX9-L, IRX10-L and IRX14-L.

3.3.1 Different composition of the xylan synthesis machinery modifies the GlcA pattern of xylan

Interestingly, the [Me]GlcA substitution patterns in *irx9*, *irx10*, and *irx14* xylan are different from the pattern exhibited by WT xylan. The even pattern of [Me]GlcA substitutions on WT xylan was consistent with that reported by Bromley et al. (2013). This pattern was revealed on the PACE analysis of GH30 xylanase products of WT xylan, which showed a predominance of $U^{[me]}X_8$ followed by $U^{[me]}X_{10}$ and $U^{[me]}X_6$, while the oligosaccharides with odd number of backbone residues were less abundant. *irx10* showed a preference for even spacing between GlcA, but $U^{[me]}X_6$ was the predominant species produced by GH30 instead of $U^{[me]}X_8$. The GH30 products of *irx9* and *irx14* xylan had no clear predominance for even or odd number of backbone residues. Also, the proportion of oligosaccharides with even to odd number of backbone residues was reduced in all mutants. Since the loss of the pattern may have been caused by the short chain length that these mutants exhibit, *irx15/15-l* was included as a control with

short xylan chains and unaltered XSC. The [Me]GlcA substitution pattern exhibited by *irx15/15-l* xylan was different to *irx9*, *irx10* and *irx14*. *irx15/15-l* xylan exhibited a clear even substitution pattern, however, the predominant oligosaccharide produced by GH30 was $U^{[me]}X_6$ which is possibly because of the shortening of the xylan length. However, *irx9*, *irx10* and *irx14* xylan exhibited a greater effect on the [Me]GlcA pattern than of *irx15/15-l*. This implies that either: the xylan in *irx9*, *irx10* and *irx14* is shorter than in *irx15/15-l*; or if the xylan size is the same, there is an additional effect on the [Me]GlcA substitution pattern. In the latter alternative, the composition of the XSC could be influencing the downstream xylan glucuronidation process.

In a recent report, Grantham et al. (2017) showed that the [Me]GlcA pattern of xylan by GUX1 changes when a xylan acetyltransferase (ESK1) is inactivated. This was shown via GH30 xylanase digestion of *esk1* mutant xylan, where the products exhibited no preference for even or odd number of backbone residues. This implies that the even spacing of [Me]GlcA produced by GUX1 is disrupted in *esk1* mutant. This mutant has nearly 50% reduction of xylan acetylation, and the main alternating pattern of acetylation is disrupted (Grantham, 2016; Xiong et al., 2013). Here, the inactivation of IRX9, IRX10 and IRX14 exhibited a change in the [Me]GlcA pattern of xylan, but this change is not attributed to a deficit in acetylation or a disruption of the acetylation pattern (section 3.2.2). Grantham et al. (2017) independently inactivated GUX1 and GUX2 in the *esk1* mutant to determine which of the glucuronyltransferases is dysregulated. *esk1 gux1* showed a non-significant 35% increase in the number of GlcA per 100 Xyl respect to *gux1*, and *esk1 gux2* (where only GUX1 is active) showed an altered even substitution pattern of xylan. In a similar manner, it would be interesting to identify which GUX is dysregulated when the XSC composition is altered. This could be achieved by analysing the [Me]GlcA pattern of xylan when inactivating a member from the XSC, and either GUX1 or GUX2.

IRX9-L overexpressors restored *irx9* short xylan to WT size, and they rescued the [Me]GlcA pattern of xylan, however, the [Me]GlcA substitutions were more frequent than WT. The XSC in the overexpressor lines has no IRX9, but the levels of IRX9-L were boosted in secondary cell walls by expressing IRX9-L under the promoter of IRX9. The xylan of the IRX9-L overexpressor lines exhibited a similar mobility to WT xylan on PACE, indicating that the xylan length is restored to WT levels when IRX9-L is overexpressed. On the other hand, the overexpressor lines showed a restoration of

the predominance for even spacing between [Me]GlcA, which implies that the even pattern of [Me]GlcA substitutions is recovered in these lines. However, the predominant frequency of spacing between [Me]GlcA in the overexpressors was different, [Me]GlcA substitutions were spaced by six Xyl instead of eight. This indicates that the secondary cell wall XSC is capable of producing the even pattern of [Me]GlcA substitutions, but with different spacing frequencies when complexed with IRX9 or IRX9-L. The overexpression of IRX9-L increases the levels of xylan, restoring its functionality (Wu et al., 2010). Here, the plant height was restored in the overexpressors, which implies that the function of xylan was restored, but by having different frequency of even-spaced [Me]GlcA to WT. This suggests that the function xylan is maintained when the frequency of spacing between [Me]GlcA is changed but the even spacings are predominant.

IRX9 (in WT) and IRX9-L (in the overexpressor lines) are able to produce fully elongated and GUX1-[Me]GlcA-patterned secondary-cell-wall xylan when expressed with IRX10 and IRX14. In agreement with this result *in vivo*, Zeng et al. (2016) published the effects on the *in vitro* activity of the asparagus XSC when interchanging its components. They showed that the XSC formed with IRX9 or IRX9-L elongates Xyl_5 with a similar efficiency when complexed with IRX10 and IRX14B. However, in the same publication they showed that when XSC has a different IRX14 homologue, the activities when complexed with IRX9 is different to when complexed with IRX9-L. This suggests that the effect of the composition of the XSC on its activity is given by more parameters than just the exchange of one component for another. Nevertheless, it is possible to conclude that IRX9-L can effectively elongate xylan to wild-type levels. This xylan can serve as a substrate to GUX1 and GUX2. Furthermore, the xylan substitution pattern and length deficiencies observed in *irx9* are due to the lack of IRX9 activity and also possibly because of a different function of IRX9-L.

3.3.2 Xylan acetylation does not explain the modified GlcA pattern in the XSC mutants

To study the effects of the XSC altered composition, the GlcA on xylan was genetically removed by inactivating GUX1 and GUX2, in addition to the respective IRX protein. Interestingly, the hydrolysis of the xylan from the *irx gux1/2* mutants with GH10 produced a higher abundance of oligosaccharides with an even number of Xyl in the backbone, which is similar to the GH10 products on the control *gux1/2*, and consistent with the pattern of *gux1/2* xylan proposed by Busse-Wicher et al. (2014). This implies that the main alternating pattern of acetylation is mostly unaffected by the modification in the XSC composition. Therefore, the effect observed in the [Me]GlcA pattern in *irx9*, *irx10* and *irx14* cannot be explained by their xylan acetylation.

A different effect was found by Grantham et al. (2017) in which the xylan [Me]GlcA pattern is altered when the ESK1-dependent Ac pattern is modified. It is possible that the XSC and ESK1 might physically interact. During the progress of this work, it was thought that the interaction between ESK1 and the XSC could have been specific for certain composition of the complex from secondary cell wall; then in *irx9*, *irx10* and *irx14* the XSC might not interact in the same way with ESK1. This could alter the acetylation pattern, which would explain their modified [Me]GlcA pattern. However, the unmodified main alternating pattern of acetylation in the *irx gux1/2* mutants suggests there may be other factors that can affect the patterning of GlcA on xylan, in addition to the acetylation pattern by ESK1.

The technique here used to analyse the pattern of acetylation in the mutant was based on a previously published methodology (Busse-Wicher et al., 2014). In disagreement, Chong et al. (2014) reported that *irx9* and *irx14* xylan acetylation is altered. The authors showed 2-D NMR spectra of DMSO-extracted xylan which showed a reduction in mono-acetylated Xyl in *irx9* and *irx14* xylan, while the acetylation in *irx10* xylan seemed to be unaltered. Here, the acetylation pattern was analysed in the absence of [Me]GlcA substitutions and through enzymatic profiling, which provides a broader scope about the pattern of Ac on xylan. The genetic elimination of [Me]GlcA, however, produces an increase in the amount of mono-substituted Ac (Busse-Wicher et al., 2014; Chong et al., 2014), possibly because the lack of [Me]GlcA produces ‘free OH’ to be acetylated. This implies that the xylan of the *irx gux1/2* mutants analysed here

may have a greater amount of mono-acetylated Xyl. However, WT and *gux1/2* xylan is acetylated mostly on every other Xyl, and therefore, the main alternating pattern of acetylation should not vary in *irx9*, *irx10* and *irx14* when inactivating GUX1 and GUX2. In addition, because the Ac substitutions on xylan are mostly two backbone residues apart, the short xylan exhibited by *irx9*, *irx10*, *irx14* and *irx15/15-l* should not be affected by the short xylan chain. However, it is uncertain that other acetylation patterns such as consecutive acetylated Xyl or di-acetylated Xyl may be affected by the composition of the XSC. Other acetylation patterns are further described later in this thesis, in Chapter 4 and Chapter 5.

The xylan chain length in the mutants seems to explain most of the differences in the [Me]GlcA pattern observed. Since the total xylan length is uncertain, other factors could possibly affect the pattern. To determine whether the effect on the pattern is only caused by length reduction it would be useful to determine the xylan length in the XSC mutants and in *irx15/15-l*. Since xylan is a linear, and each polymer carries a single reducing-end-oligosaccharide (REO) (Li, 2011), a close estimation could be achieved by calculating the stoichiometric ratio of REO:Xyl. For this, the number of Xyl per xylan chain can be measured by DASH using GH10 xylanase as described by Grantham (2016).

In the XSC mutants' xylan the production of oligosaccharides with odd number of backbone residues by GH30 was increased. Since the odd spacing between [Me]GlcA are characteristic of GUX2 activity, the region of xylan that is reduced in the mutants could be decorated by GUX1. To investigate whether [Me]GlcA decorations by GUX1 or GUX2 are affected in the XSC mutants, it could be useful to analyse the pattern in when inactivating GUX1 or GUX2.

Chapter 4: The acetylation pattern in the compatible and incompatible domains of xylan

4.1 Introduction

Xylan of eudicots has two domains with different [Me]GlcA substitution patterns (Bromley et al., 2013). The xylan compatible domain can interact with hydrophilic surfaces of cellulose, whereas the xylan incompatible domain is thought to bridge these xylanocellulose interactions (Grantham et al., 2017; Simmons et al., 2016). In this model, the xylan substitution patterns are fundamental for the function of each domain. In the compatible domain, the xylan backbone adopts a two-fold-screw configuration, with [Me]GlcA substitutions facing away from the cellulose. This is possible because the [Me]GlcA are separated by even number of backbone Xyl. In contrast, the distance between the GlcA substitutions in the incompatible domain has no preference for even or odd number of backbone residues. Ac decorations on xylan are much more frequent, substituting every other backbone Xyl (see Chapter 1, section 1.4.4.2, page 33) (Busse-Wicher et al., 2014; Chong et al., 2014).

The compatible and incompatible domains are thought to be covalently linked, but the [Me]GlcA modifications on them are product of two enzymatic activities (Bromley et al., 2013). GUX1 and GUX2 decorate the compatible and incompatible domains respectively. The inactivation of GUX1 or GUX2 depletes [Me]GlcA in the respective xylan domain, while the enzyme that remains active cannot compensate. It is not understood how GUX1 and GUX2 recognise the specific region of xylan that they decorate, and how they generate their specific [Me]GlcA pattern. The inactivation of the xylan acetyltransferase ESK1 produces a dysregulation of the activity of GUX1 and possibly GUX2 (Grantham et al., 2017). In *esk1* xylan, the region that is decorated by GUX1 is extended, which suggests that GUX1 and GUX2 act on portions of xylan with different acetylation environments. In this hypothesis, the xylan compatible and incompatible domains are first determined by their acetylation, and then glucuronidated. In the compatible domain, Ac and [Me]GlcA substitution patterns are possibly ‘in frame’, maintaining the even number of Xyl between both types of modification. However, the acetylation patterns of other xylan regions have not been reported. The characterisation of the acetylation patterns in the specific regions of

xylan that are substituted by GUX1 and GUX2 could provide information about the specificity determinant for each xylan glucuronyltransferase and perhaps a possible function of di-acetylated Xyl. Investigating the Ac on xylan decorated by GUX2 would support the hypothesis that incompatible xylan has a different pattern around [Me]GlcA substitutions.

The aim of this chapter was to describe the acetylation in the compatible and the incompatible domains of xylan from secondary cell walls. In order to do this, *gux1* and *gux2* xylan was studied because they each lack [Me]GlcA decorations on one of the two domains, allowing fragments derived from each domain to be studied alone. To hydrolyse the xylan, two xylanases were first evaluated, determining the complexity of their products, and then the most suitable was chosen. The relative positions of the Ac respect to the [Me]GlcA substitutions on the products were determined by tandem mass-spectrometry.

4.2 Results

The acetylation pattern was assessed by relating it to the [Me]GlcA present in the compatible or incompatible domains. The xylan of *gux1* has GlcA only in the incompatible domain, and *gux2* has [Me]GlcA only in the compatible domain (Bromley et al., 2013). The [Me]GlcA substitutions in xylan from *gux1* and *gux2* were therefore used to identify the origin of the digestion products. It was hypothesised that Ac near to [Me]GlcA in the compatible domain might be spaced an even number of residues from the [Me]GlcA.

The later MS analysis of the xylan hydrolysis products requires the sample composition to be as simple as possible, because isobaric oligosaccharides would make interpretation of spectra difficult. This is because the MS fragmentation spectra of isobaric oligosaccharides would overlap (see Chapter 1, section 1.5.3). For this, GH30 and GH10 products were first analysed by PACE, but only one was selected for the MS analysis, based on production of the simplest acetylated glucuronoxylan oligosaccharides. These two enzymes were chosen because both can digest acetylated xylan. However, they have different specificities. GH30 requires a [Me]GlcA (or Ac with less affinity) in the -2 subsite, while GH10 affinity is more promiscuous and does not require a specific xylan substitution to hydrolyse the backbone (see Chapter 1, section 1.5.2).

4.2.1 GH10 and GH30 hydrolysis evaluation on *gux1* and *gux2* xylan

Ideally, the xylanase that produces the largest proportion of (methyl)glucuronidated to non-glucuronidated products is most suitable for the later MS analysis. In order to do this, the GH30 and GH10 products of *gux1* and *gux2* xylan were compared on PACE. *gux1* and *gux2* plants were grown and their basal stems were used to prepare AIR. The material was sequentially depectinated with ammonium oxalate and delignified with peracetic acid, and then the acetylated xylan was extracted with DMSO. The xylan was hydrolysed with GH10 or GH30, and then the enzymes were inactivated with heat. The reaction products were divided, and some were treated with alkali to remove all *O*-acetyl groups. The acetylated and de-acetylated products were analysed by PACE (Figure 4.1 panel A and panel B respectively).

GH10 digested a higher proportion of *gux1* xylan than GH30 (panel A) since the signal produced by its products on the gel is higher (for PACE analysis see Chapter 1, section 1.5.3). On the other hand, it was not clear whether GH10 or GH30 hydrolysed *gux2* xylan most efficiently.

For the samples treated with alkali (panel B), since all Ac were removed, oligosaccharides that co-migrated with the Xyl standard lacked [Me]GlcA substitutions. The six *gux1* GH10 products with highest abundance co-migrated with X_1 to X_6 of the Xyl standard, therefore, these oligosaccharides did not carry [Me]GlcA. Also some *gux1* GH30 products co-migrated with the standards, however, more oligosaccharides did not co-migrate with the standards compared to *gux1* GH10 products. This indicates that on *gux1* xylan, GH30 produced a higher proportion of oligosaccharides that carried a [Me]GlcA to the ones that lacked [Me]GlcA, compared to GH10. For *gux2* xylan, GH10 hydrolysis mostly produced oligosaccharides that co-migrated with X_1 to X_5, and X_6 was not predominant. GH30 hydrolysis of *gux2* xylan mostly produced X_2 and X_4. Several other oligosaccharides were also produced which did and did not co-migrate with the standards with no clear dominance. This indicates that on *gux2* xylan, GH30 produced a higher proportion of oligosaccharides that carried a [Me]GlcA to the ones that lacked [Me]GlcA, compared to GH10. This is consistent with the activity of GH30, which has higher affinity for [Me]GlcA over Ac at the subsite -2 (Busse-Wicher et al., 2014). The non-glucuronidated secondary products are formed when acetylated xylooligosaccharides carrying a single [Me]GlcA (initial GH30 products) are further

hydrolysed by GH30. Therefore, the presence of GH30 non-glucuronidated products indicates that the hydrolysis with GH30 was long. However, for the purpose of this chapter, GH30 was the most suitable enzyme to use for the MS analysis, because most of the oligosaccharides that it produced were substituted with [Me]GlcA.

The GH10 hydrolysis of *gux2* xylan (which is decorated by GUX1) produced more (methyl)glucuronidated products than the GH10 hydrolysis of *gux1* xylan. This is consistent with the proportion of xylan in these mutants that is (methyl)glucuronidated (70% and 30% respectively; Bromley et al., 2013). Therefore, this suggests that GH10 had more accessibility to the non-[Me]GlcA-substituted portions of xylan. On the other hand, it is unclear whether *gux1* or *gux2* produced more (methyl)glucuronidated products when hydrolysed with GH30. However, GH30 produced a larger amount of smaller oligosaccharides on *gux1* (red arrow) and larger oligosaccharides on *gux2* (blue arrows). This indicates that [Me]GlcA substitutions in *gux2* xylan (by GUX1) were more distant than in *gux1* xylan (by GUX2). This is consistent with the reported predominant larger spacing between [Me]GlcA substitutions by GUX1 compared to [Me]GlcA by GUX2 (Bromley et al., 2013).

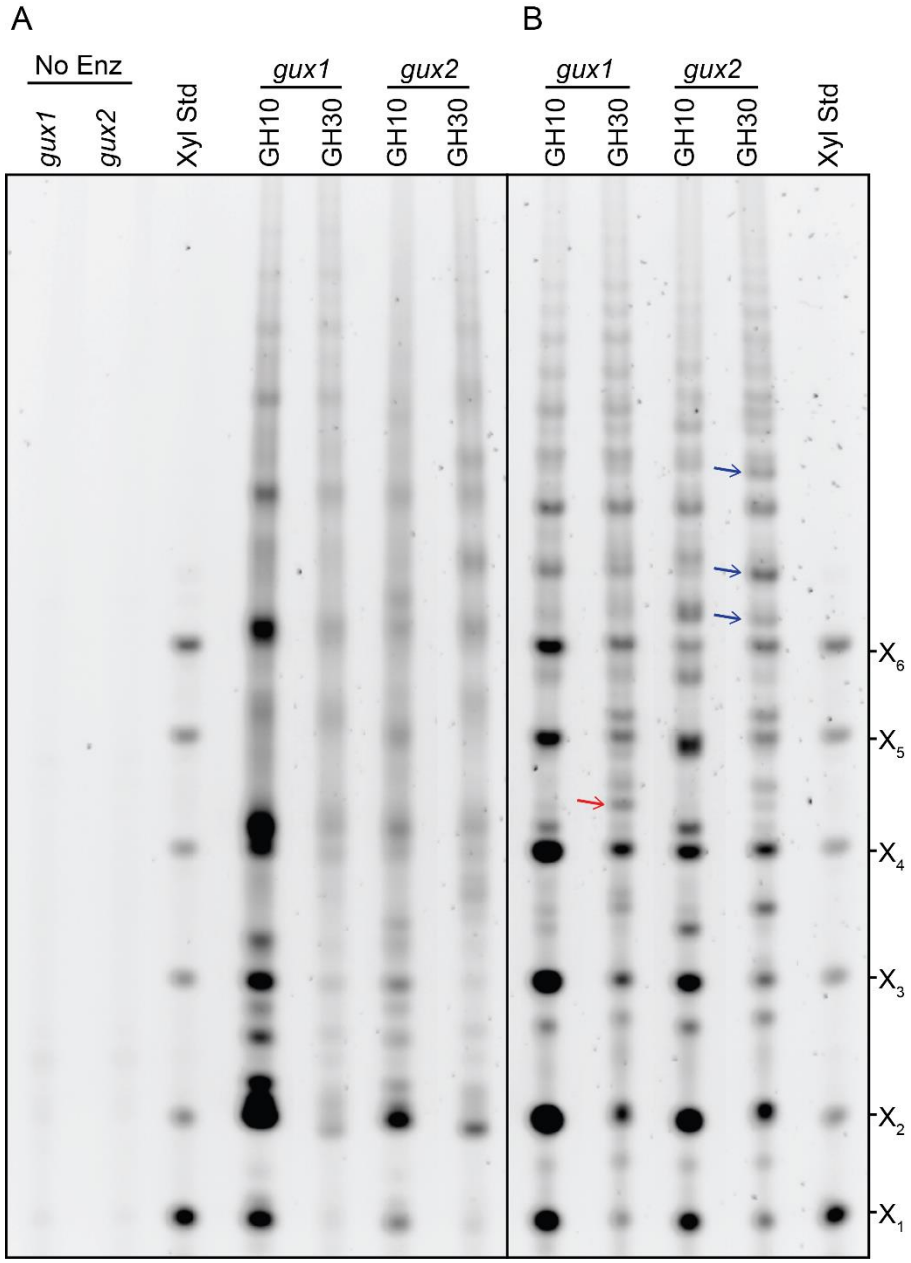


Figure 4.1: Comparison of EcGH30 and GH10 hydrolysis products of *gux1* and *gux2* xylan.
Xylan was extracted with DMSO and incubated for 16 h at 21 °C. The enzymes were inactivated with heat, and the products were de-acetylated (B) or not (A) with alkali. The samples were derivatised with ANTS and resolved on a 10% polyacrylamide PACE gel. Blue and red arrows indicate differential species obtained from *gux1* and *gux2* xylan when digested with GH30. Lanes of this gel were rearranged to ease interpretation.

4.2.2 Description of the acetylation in the compatible and incompatible domains

To analyse the acetylation patterns in the xylan compatible and incompatible domains, *gux1* and *gux2* xylan was digested with GH30 and the oligosaccharides were analysed by MS. However, the more isobaric isomers a sample has, the more complex it is to analyse by mass-spectrometry. For example, the oligosaccharide products might have identical number of Xyl, but differently positioned Ac. The GH30 oligosaccharide products with an undecorated NRE-terminal Xyl followed by a mono-acetylated Xyl (XX^{a}~) are isomeric with the ones with an inverted order ($X^{a}X$~), and the MS/MS fragmentation spectra would overlap. To reduce the amount of complexity of the sample produced by GH30, any unsubstituted NRE Xyl were enzymatically removed with GH3. This enzyme removes all unsubstituted Xyl from the NRE until it reaches an acetylated residue (see Chapter 1, Figure 1.31). After this GH30/GH3 treatment, the products have a substituted Xyl at the NRE (presumably an Ac). Therefore, the distance between the [Me]GlcA (at subsite -2) and the acetyl substitution at the NRE could be determined, because it equals the number of Xyl in the backbone minus 2. Since of the non-glucuronidated GH30 products were highly abundant and they could interfere with the MS detection of low abundance oligosaccharides, the hydrolysed oligosaccharides were purified by solid-phase extraction with a C18 column. Then the samples were labelled with 2AA at the RE, and the labelling solution was removed from the sample using a GlycoClean S column. The samples were fractionated by HILIC to separate isomers, and analysed by MALDI-ToF mass-spectrometry.

4.2.2.1 Acetylation in the compatible domain

The mass-spectra obtained of xylan oligosaccharides from *gux2* mutant plants is shown in Figure 4.2. Since *gux2* lacks [Me]GlcA in the incompatible domain of xylan, the charged oligosaccharides from *gux2* are derived from the compatible domain. In panel A the spectrum of the sample used as input for HILIC separation is shown. The labelled peaks correspond to the sodiated adducts of the oligosaccharides detected ($[M + Na]^+$). The predicted oligosaccharide species are colour-labelled depending on their composition, to aid interpretation. The species with no [Me]GlcA are labelled in cyan. The other species had one [Me]GlcA; the species in green had even number of pentosyl residues (Pent), and the species in red had odd number of Pent. As expected, non-[Me]GlcA modified species had very low abundance, as they were removed by the C18 column. There were both MeGlcA and -14 Da GlcA forms of most oligosaccharides, and mostly MeGlcA was expected because of its natural abundance in *Arabidopsis* secondary-cell-wall xylan (Bromley et al., 2013; Chong et al., 2014). The predominant species had an even number of Pent, indicating they had even number of backbone Xyl between the [Me]GlcA and the Ac on the NRE. This supports the hypothesis that the [Me]GlcA pattern in the compatible domain is 'in frame' with the alternating acetylation pattern. However, low abundance structures with odd number of Pent were also detected.

To determine the position of the Ac substitutions in relationship to the [Me]GlcA, the molecular ions were selected from the HILIC fractions (based on predicted compositions that included the mass of [Me]GlcA) and analysed by MALDI TOF-TOF MS/MS. The structures found are shown in Figure 4.2 panels B – N. The structures shown were determined by following the Y and B ion series for each MS/MS spectrum (Domon and Costello, 1988). The main oligosaccharides were all compatible, just varying in the position of mono-and di-acetylated Xyl, presence or absence of methylation on the GlcA, and Ac on the Xyl carrying a [Me]GlcA. Some small amount of exceptions were also found. Thus, the method worked, and it is supporting the view that the [Me]GlcA and Ac are largely compatible in spacing. To consider the variety in the substitution patterns on *gux2* xylan, the structure of GH30/GH3 products is next discussed individually.

The species with highest abundance was $Pent_4Ac_3U^{me}$ (*m/z* = 1006), of which two isomeric structures were found. The compatibly patterned structure $X^{aa}XU^{me,a}X$ (panel D) predominated over $X^aX^aU^{me,a}X$ (panel E) because it had a higher signal on the electropherogram (data not shown). In both structures the glycosidic bond cleavage $Y_{2\alpha}$ (*m/z* = 816) was diagnostic of the MeGlcA substitution. Y_1 (*m/z* = 294) was diagnostic of the unsubstituted Xyl at the RE (characteristic of GH30 products) linked to 2AA. The difference between Y_1 and Y_2 (*m/z* = 658) indicates that the following Xyl towards the NRE carried a MeGlcA and an Ac. In structure D, Y_3 (*m/z* = 790) indicates that the next Xyl towards the NRE was unsubstituted, and the difference between Y_3 and the total mass denotes that the NRE Xyl was di-acetylated. Instead, in structure E, Y_3 (*m/z* = 832) indicates that the two NRE Xyl were mono-acetylated.

Other structures with four backbone Xyl, $X^aXU^{me,a}X$ (*m/z* = 964, panel B) and $X^{aa}XU^aX$ (*m/z* = 992, panel C) also had two acetylated Xyl separated by an unsubstituted Xyl. In structure B, Y_1, Y_2 and Y_3 were similar to structure D described above. In structure C, $Y_{2\alpha}$ (*m/z* = 816) was diagnostic of a non-methylated GlcA substitution, and the difference between Y_1 and Y_2 (*m/z* = 644) indicates that the second Xyl from the RE carried the GlcA and an Ac. Then, Y_3 (*m/z* = 776) denotes that the following Xyl towards the NRE was unsubstituted, and that the last one was di-acetylated.

The species with the second highest abundance was $X^aXX^aXX^aXU^{me,a}X$ (*m/z* = 1576, panel N) in which the Y-ion series indicates alternating acetylated Xyl. The mass differences Y_3-Y_2 and Y_4-Y_3 denote the consecutive residues X^aX~ next to the MeGlcA-substituted Xyl. $Y_{4\text{-}7}$ ions exhibited a similar behaviour, indicating alternating mono-acetylated Xyl. This is supported by a similar mas difference exhibited by the B-ion series, e.g. B_4-B_3 equals the mass of Xyl (132 Da) and B_5-B_4 is the mass of a mono-acetylated Xyl (174 Da). The species $X^aXX^aXX^aXU^{me}X$ (*m/z* = 1534, panel M) also had eight backbone residues and alternating substituted Xyl, but the MeGlcA-modified Xyl was not acetylated. This is denoted by the Y_2-Y_1 difference of 322 Da. Structures with *m/z* = 1618 and *m/z* = 1660 (panel A, $Pent_8Ac_5\ U^{me}$ and $Pent_8Ac_6U^{me}$, respectively) also had eight backbone Xyl according to their MS/MS fragmentation spectra (data not shown), however, the isomeric structures were not resolved by HILIC, precluding structural interpretation by MS/MS.

The two predominant structures with six backbone residues exhibited alternating acetylated Xyl. The modified Xyl on $X^{a}XX^{a}XU^{me,a}X$ (*m/z* = 1270, panel I) were mono-acetylated, while *m/z* = 1312 (panels J and K) appeared to be a mixture of $X^{a}XX^{aa}XU^{me,a}X$ and $X^{aa}XX^{a}XU^{me,a}X$, respectively. Another species with six backbone residues and lower abundance, $X^{a}XX^{a}XU^{a}X$ (*m/z* = 1256, panel H) also showed alternated substituted Xyl. Interestingly, the structure $X^{a}XXXU^{me,a}X$ (*m/z* = 1228, panel G) showed a three-Xyl gap between the decorated Xyl.

$X^{aa}XP^{a}X$ (*m/z* = 1124, panel F) had an odd number of pentoses in its molecular mass. The difference between that total molecular mass and $Y_{2\alpha}$ (*m/z* = 816) corresponds to the mass of a GlcA-substituted Xyl (308 Da), indicating a Pent modification on the GlcA. This is supported by the absence of a peak (*m/z* = 948) corresponding to the loss of an unsubstituted GlcA. The link between the GlcA and this pentose was not determined, but it was drawn as a 1,2-linked pentopyranose (PUX structure), based on the data of Mortimer et al. (2015) and Chong et al. (2015). This structure possibly originated from a low content of primary-cell-wall xylan in the AIR preparation (see Chapter 5).

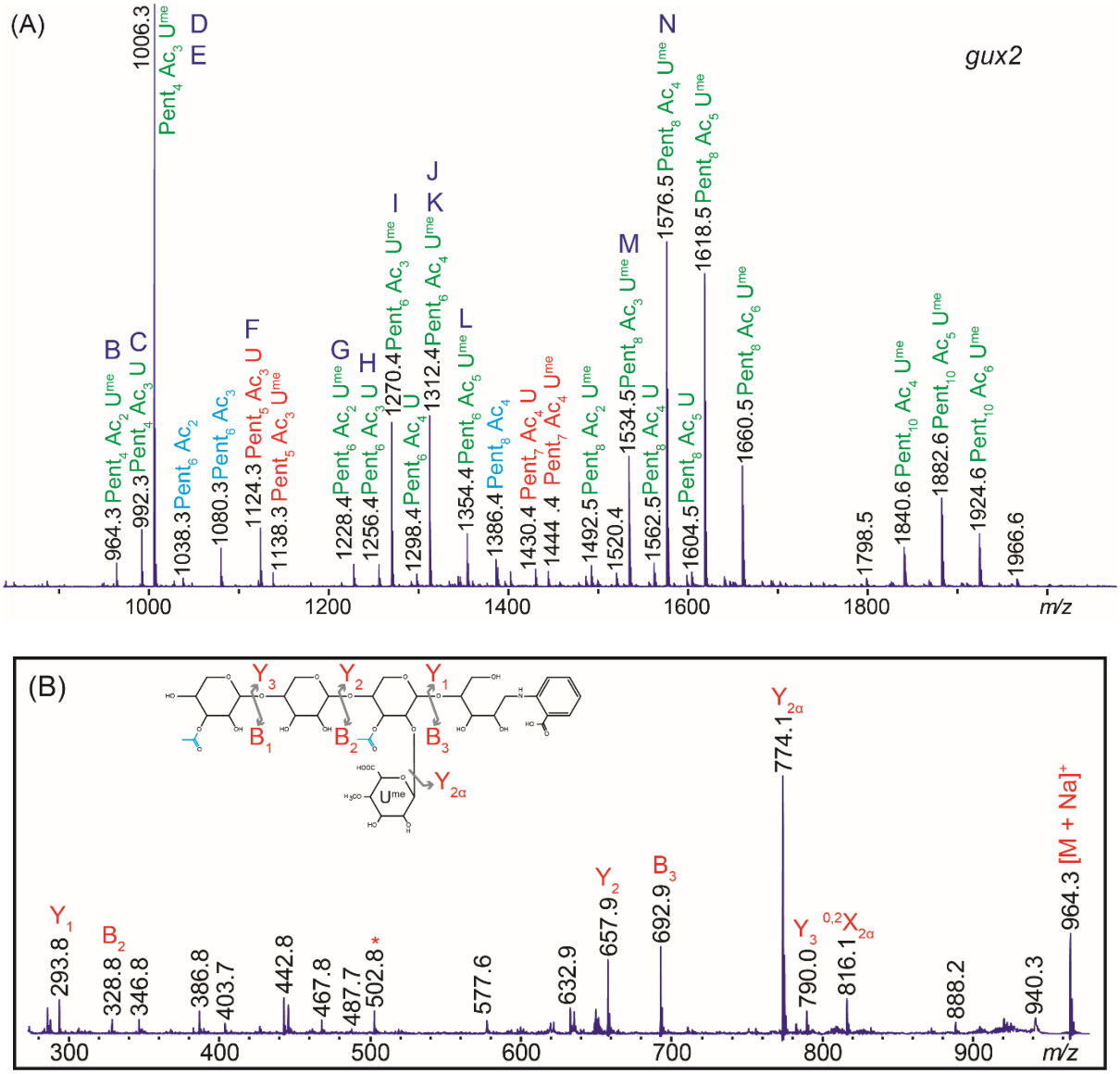
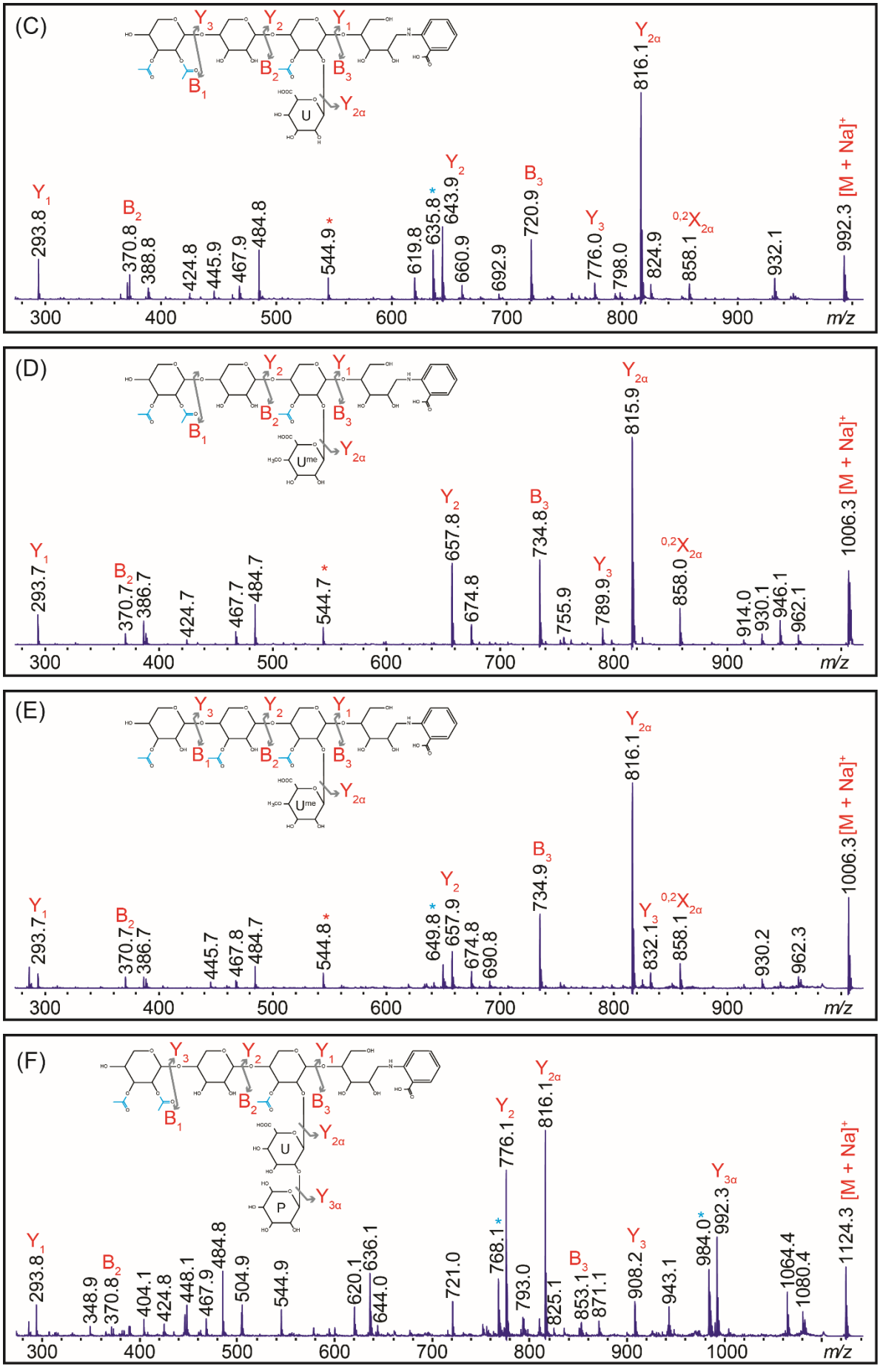
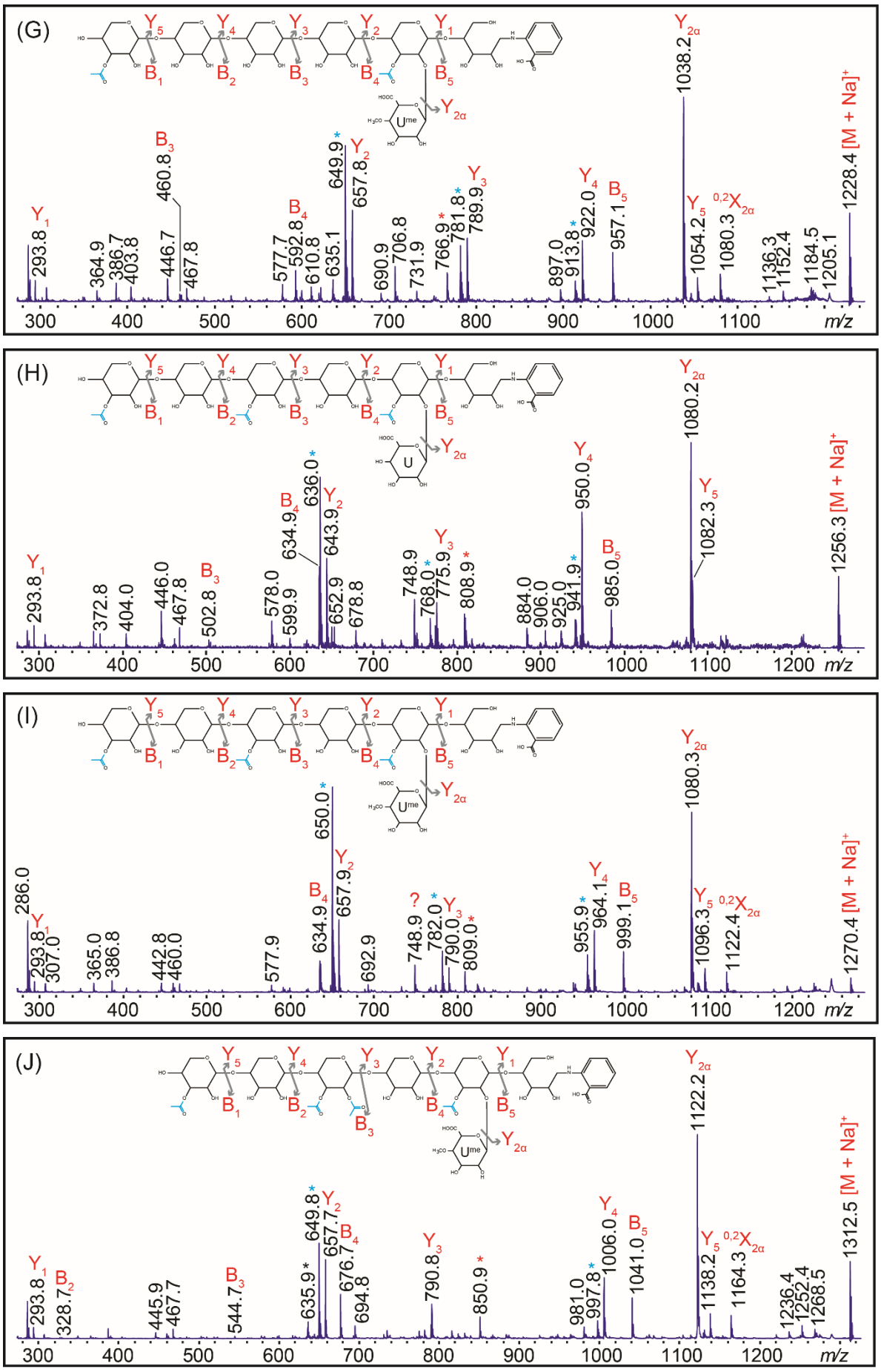
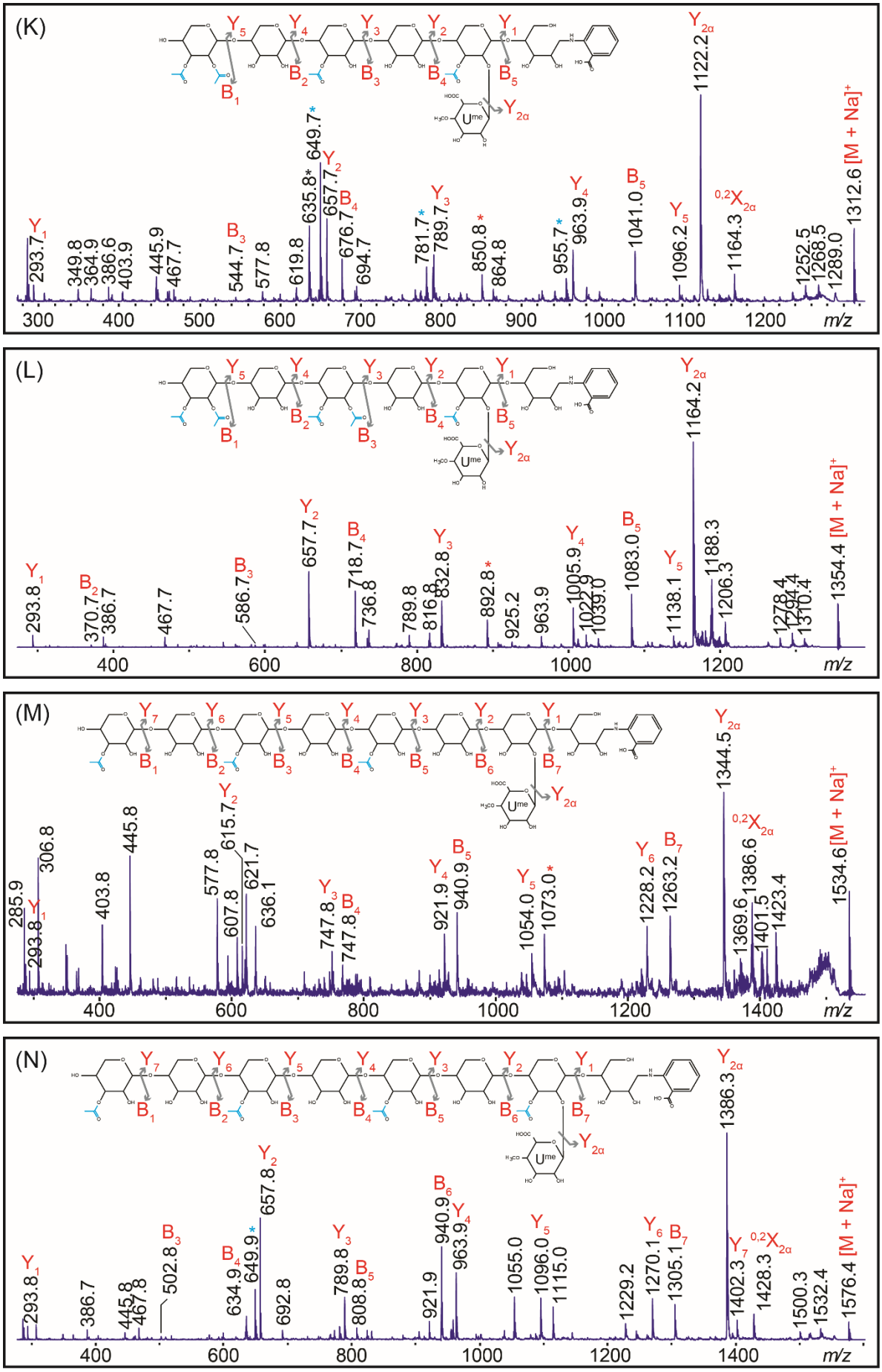


Figure 4.2: Mass spectrometric analysis of *gux2* xylan acetylation.
Mass spectra of DMSO extracted xylan digested sequentially with EcGH30 and GH3. The products were purified with a C18 column, labelled with 2AA and purified with a GlycoClean S cartridge. The spectrum in panel A was acquired in reflector mode and the peaks correspond to sodiated adducts of the predicted oligosaccharides; structures with no [Me]GlcA decorations (cyan), [Me]GlcA-modified structures with even number of Pent (green), and [Me]GlcA-modified structures with odd number of Pent (red) are indicated. The products were then separated by HILIC, and ToF/ToF MS/MS spectra were acquired. The determined structures in increasing mass ([M + Na]$^+$) and their spectra are shown in panels B – N. The linkages of the Ac on the mono-acetylated backbone residues were not determined and were drawn as O-3-linked. Asterisks: black – unidentified peak; blue – Y ion minus 8 Da for an unknown reason; green – Y ion minus 24 Da for an unknown reason; red – B ion with a loss of [Me]GlcA for a unknown reason; pink – Y or B ion with a loss of Pent and GlcA for an unknown reason. (continued).


(C)
Y1 293.8
B2 370.8
388.8
424.8
445.9
467.9
484.8
544.9 *
619.8
635.8 *
643.9
Y2
660.9
692.9
B3 720.9
Y3 776.0
798.0
824.9
Y2α 816.1
0,2X2α 858.1
932.1
992.3 [M + Na]+
m/z
(D)
Y1 293.7
B2 370.7
386.7
424.7
467.7
484.7
544.7 *
Y2 657.8
674.8
B3 734.8
755.9
Y3 789.9
Y2α 815.9
0,2X2α 858.0
914.0
930.1
946.1
962.1
1006.3 [M + Na]+
m/z
(E)
Y1 293.7
B2 370.7
386.7
445.7
467.8
484.7
544.8 *
649.8 *
Y2 657.9
674.8
690.8
B3 734.9
Y3 832.1
Y2α 816.1
0,2X2α 858.1
930.2
962.3
1006.3 [M + Na]+
m/z
(F)
Y1 293.8
B2 348.9
370.8
404.1
424.8
448.1
467.9
484.8
504.9
544.9
620.1
636.1
644.0
721.0
768.1 *
Y2 776.1
793.0
B3 825.1
853.1
871.1
Y3 908.2
943.1
984.0 *
Y3α 992.3
Y2α 816.1
1064.4
1080.4
1124.3 [M + Na]+
m/z


(G)
(H)
(I)
(J)
m/z
[M + Na]+


(K)
(L)
(M)
(N)
m/z
[M + Na]+

4.2.2.2 Acetylation in the incompatible domain

To study the acetylation in the incompatible domain, *gux1* xylan was also analysed following the procedure outlined above. The spectra are shown in Figure 4.3. In panel A the spectrum of the sample that was used as input for HILIC is shown, and the same composition-colour-coding was used as in Figure 4.2. All the identified peaks detected correspond to sodiated adducts of their predicted masses ($[M + Na]^+$). Most non-[Me]GlcA-modified species were removed by the C18 column, however, *gux1* has a larger portion of xylan without [Me]GlcA modifications than *gux2*, and therefore it was expected to find a higher abundance of non-acidic oligosaccharides. The two predominant species had even number of Pent, however, several species with odd number of Pent were also highly represented. The HILIC and TOF-TOF MS/MS analysis was carried out as described above, and the spectra are shown in Figure 4.3 panels B – T. In all structures determined, the second backbone Xyl from the RE was 2-*O*-[Me]GlcA modified and 3-*O*-acetylated.

The species with highest abundance was *m/z* = 1006 (panel C) for which only one structure was found; $X^{aa}XU^{me,a}X$. Its non-methylated version ($X^{aa}XU^{a}X$, panel B) was also observed with a lower abundance.

The second species with highest abundance was *m/z* = 1312 ($Pent_6Ac_4U^{me}$). Four isomeric structures with six backbone Xyl were found for this mass: $X^{a}XX^{aa}XU^{me,a}X$, $X^{a}X^{a}X^{a}XU^{me,a}X$, $X^{a}X^{a}XX^{a}U^{me,a}X$ and $X^{a}XX^{a}X^{a}U^{me,a}X$ (panels L, M, N and O, respectively). The structure L had alternating substituted Xyl, including a 2,3-di-acetylated Xyl, determined by the difference between Y_3 (*m/z* = 790) and Y_4 (*m/z* = 1006). The structures in panels M, N and O had consecutive mono-acetylated Xyl, with no X^{aa} residues.

Four structures that differed in mass (*m/z* = 1270, 1298, 1354 and 1402) had six backbone residues: $X^{a}XX^{a}XU^{me,a}X$ (panel J), $X^{a}XX^{aa}XU^{a}X$ (panel K) $X^{a}X^{a}X^{a}X^{a}U^{me,a}X$ (panel P) $NXX^{aa}XU^{me,a}X$ (panel Q, where N is a Pent-substituted Xyl), respectively. In structure P all substituted Xyl were consecutively located. In contrast, in structures J, K and Q, the substituted Xyl were alternated with non-substituted Xyl. One of the backbone Xyl in structure Q had a pentosyl side-chain, which is an unreported substitution for secondary-cell-wall xylan. This modification at the NRE was insensitive to GH3 xylosidase, indicating it is not a backbone Xyl.

A pentosyl side-chain was also found in structure $N^{a}X^{a}U^{me,a}X$ (*m/z* = 1138, panel E). In this structure, the NRE Xyl carried an Ac and a Pent of unknown identity. Since GH3 was used in this digestion, the products should not possess unsubstituted backbone Xyl at the NRE. However, Y_4 (*m/z* = 1006) denoted the absence of Ac on the last pentose of the chain. Therefore, a pentosyl sidechain is possible. All modified backbone residues in this structure were consecutively arranged.

Structures $X^{a}X^{a}X^{a}U^{a}X$ (panel F) and $X^{a}X^{a}X^{a}U^{me,a}X$ (panel G) had five backbone Xyl. Both structures had three consecutive mono-acetylated Xyl from the NRE, followed by a fourth acetylated Xyl which also carried a GlcA or MeGlcA respectively. However, structure F was very close to the noise, while structure G was the fifth with highest intensity of the sample.

Two structures were determined with an *m/z* = 1222, $X^{a}X^{a}X^{aa}U^{me,a}X$ (panel H) and $X^{a}X^{aa}X^{a}U^{me,a}X$ (panel I). Both structures had five backbone Xyl, of which four were consecutively acetylated. In both structures there was a 2,3-di-acetylated Xyl, but the position of this residue differed between structures. The difference in the position of the di-acetylated Xyl is denoted by the differences Y_3-Y_2 and Y_4-Y_3 which were interchanged between structures.

The structures determined with the longest backbone were $X^{a}X^{a}X^{a}X^{a}XU^{me,a}X$, $X^{a}X^{a}X^{a}XX^{a}U^{me,a}X$ (panels R and S respectively; both *m/z* = 1486) and $X^{a}X^{a}X^{a}X^{a}X^{a}U^{me,a}X$ panel T (*m/z* = 1528), with seven backbone Xyl. Structure T exhibited five consecutive mono-acetylated Xyl from the NRE, followed by $U^{me,a}$. Structures R and S, lacked the fourth or third Ac from the NRE, respectively. Additional structures for *m/z* = 1486 were also detected but not resolved by HILIC.

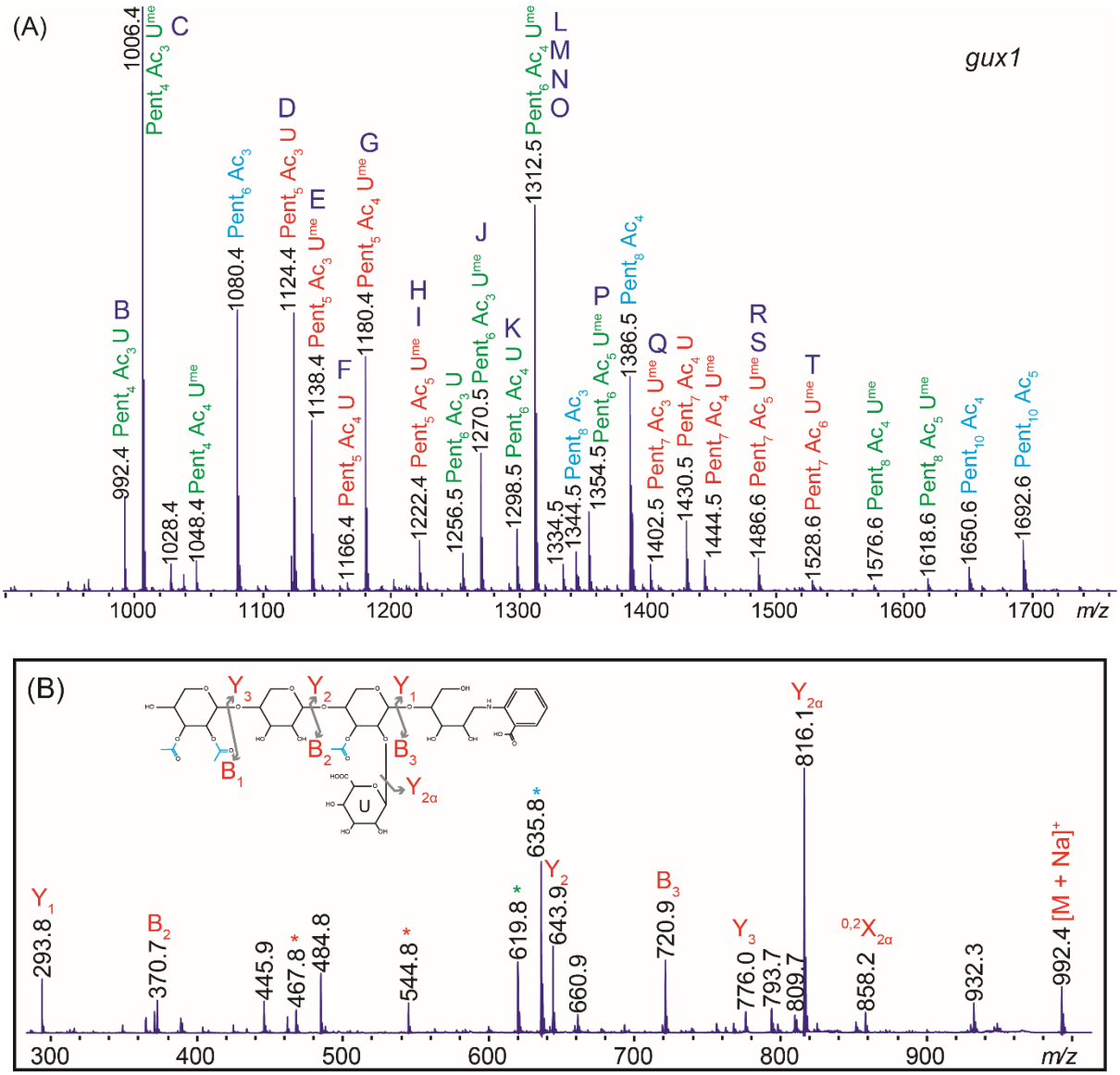
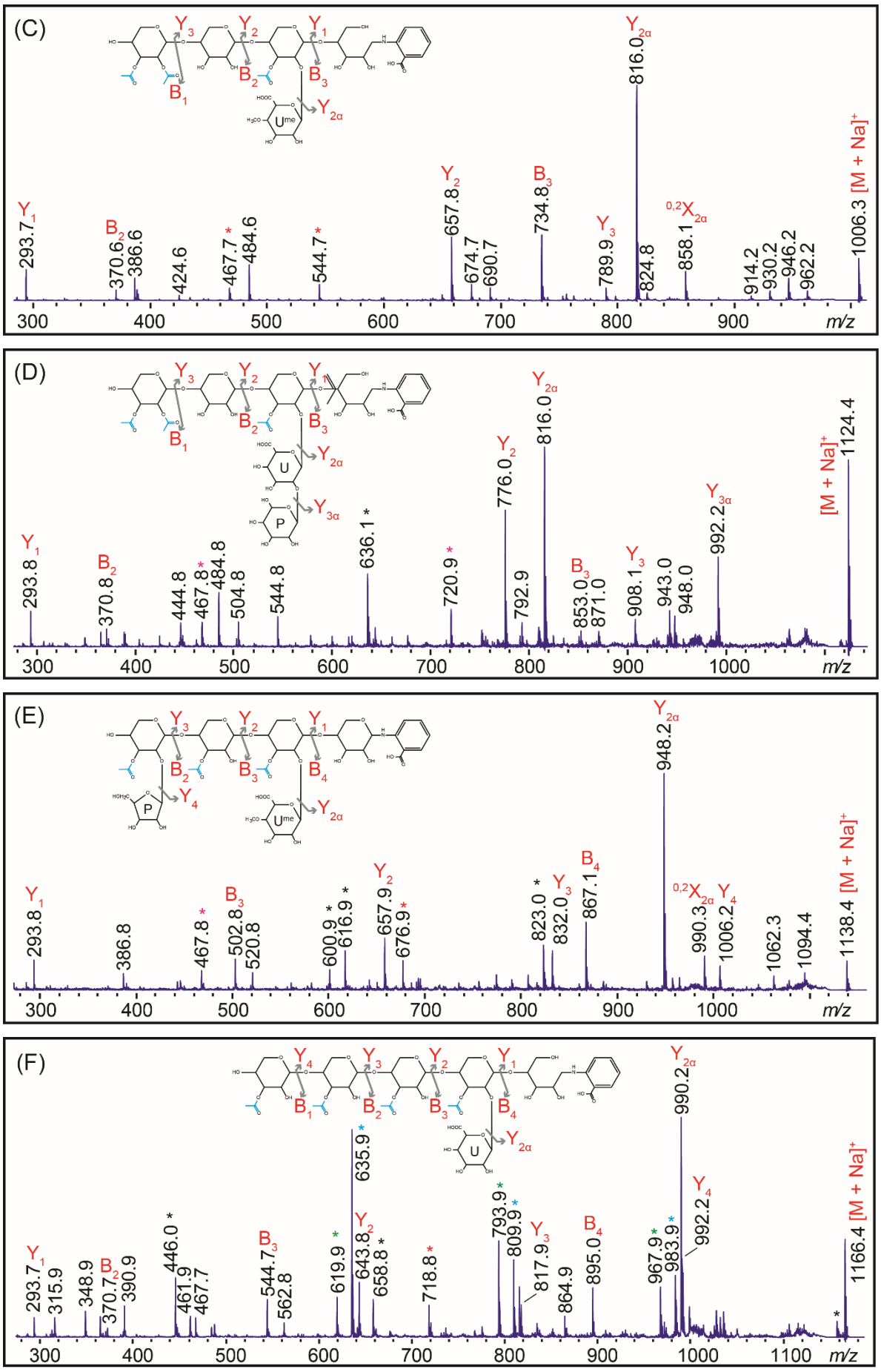
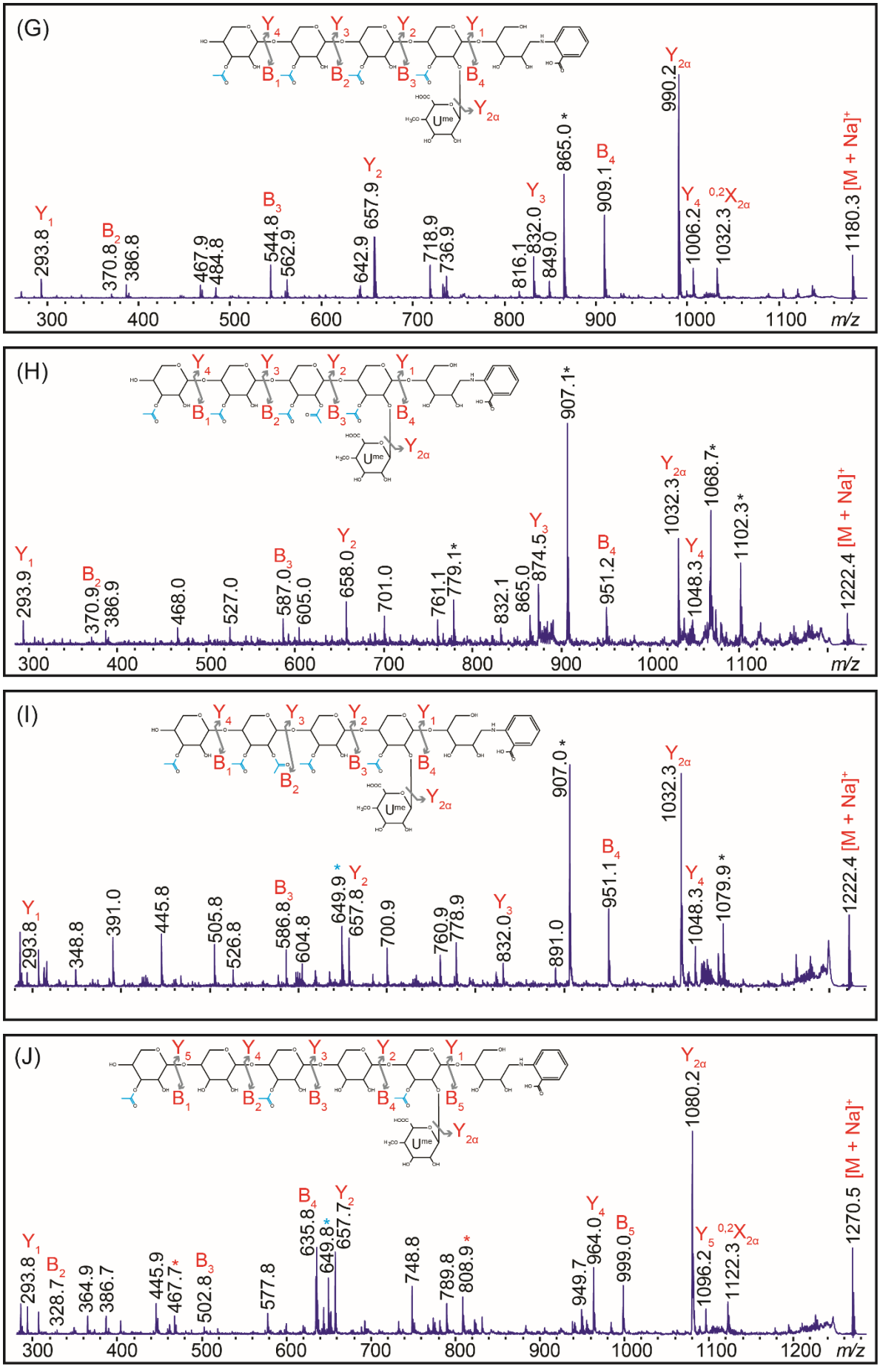
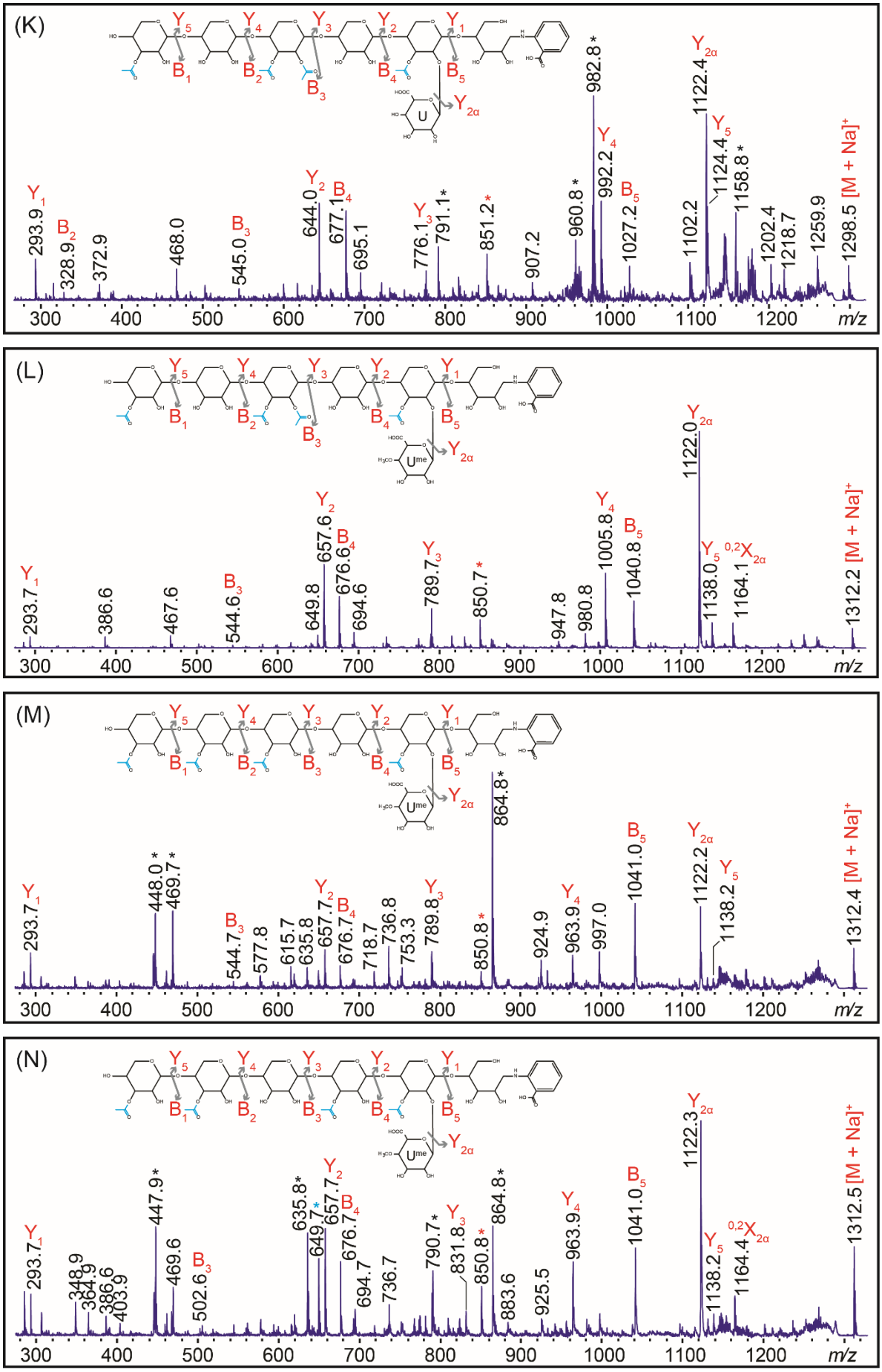
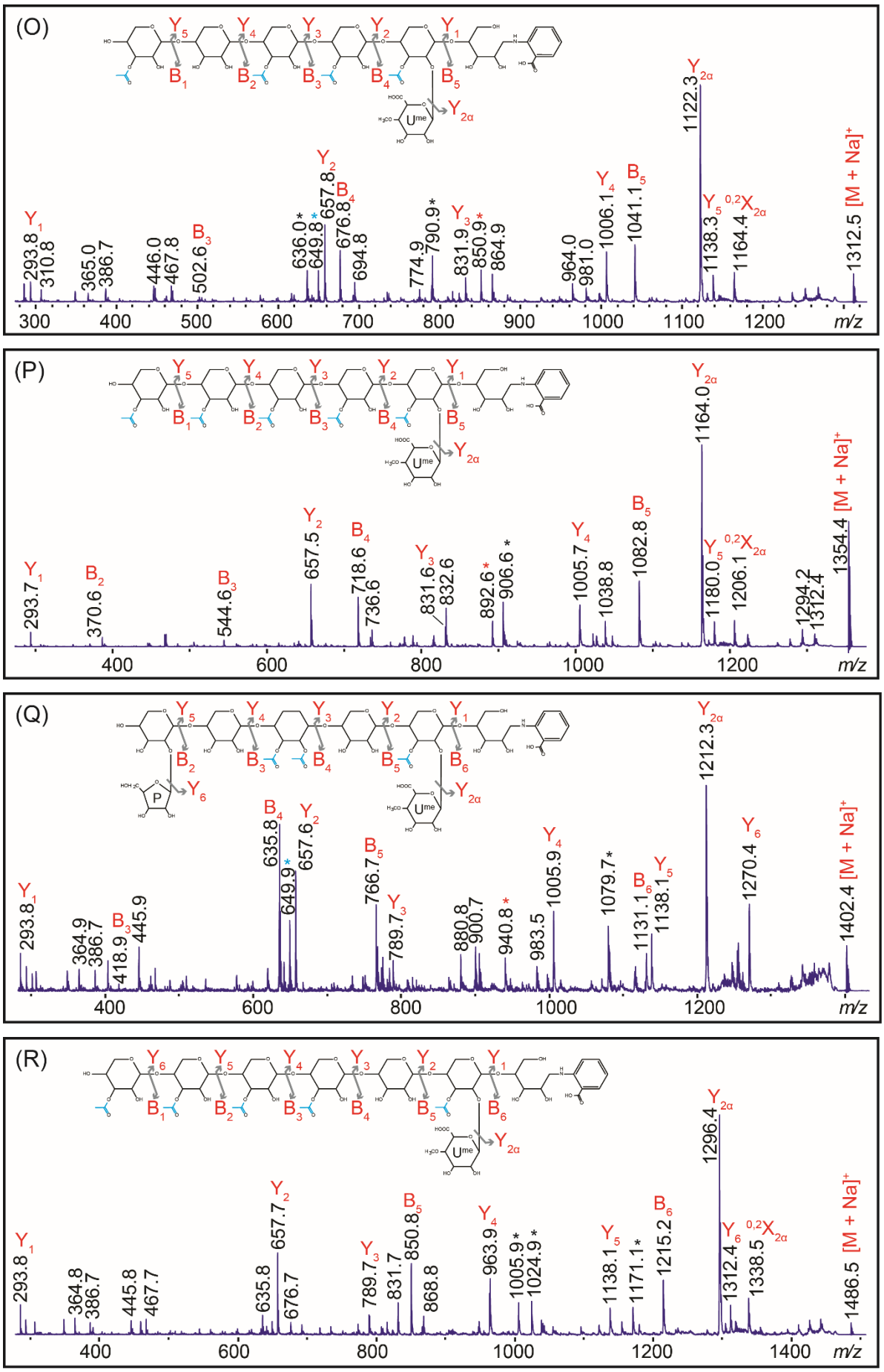


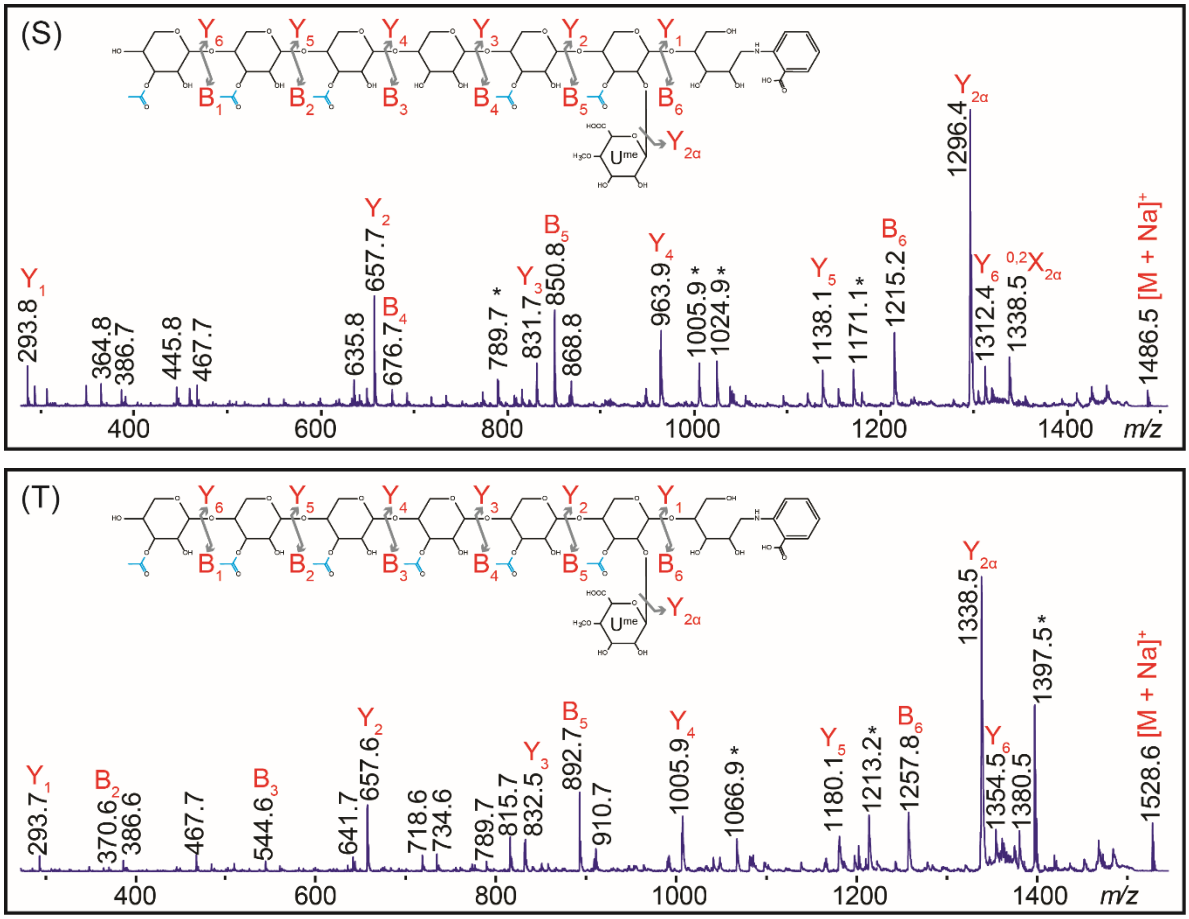
Figure 4.3: Mass-spectrometric analysis of *gux1* xylan acetylation.
Mass-spectra of DMSO extracted xylan digested sequentially with EcGH30 and GH3. The products were purified with a C18 column, labelled with 2AA and purified with a GlycoClean S cartridge. The spectrum in panel A was acquired in reflector mode and the peaks correspond to sodiated adducts of the predicted oligosaccharides; structures with no [Me]GlcA decorations (cyan), [Me]GlcA-modified structures with even number of Pent (green), and [Me]GlcA-modified structures with odd number of Pent (red) are indicated. The products were then separated by HILIC, and ToF/ToF MS/MS spectra were acquired. The determined structures in increasing mass ($[M + Na]^+$) and their spectra are shown in panels B – T. The linkages of the Ac on the mono-acetylated backbone residues were not determined and were drawn as O-3-linked. The identity and linkage of the pentosyl modifications on the backbone were also not determined, but were drawn as 1-2-linked pentofuranosyl residues. Asterisks: black – unidentified peak; blue – Y ion minus 8 Da for an unknown reason; green – Y ion minus 24 Da for an unknown reason; red – B ion with a loss of [Me]GlcA for a unknown reason; pink – Y or B ion with a loss of Pent and GlcA for an unknown reason. (continued).


(C)
Y1 293.7
B2 370.6 386.6
424.6
467.7 * 484.6
544.7 *
Y2 657.8 674.7 690.7
B3 734.8
Y3 789.9
Y2α 816.0 824.8
0,2X2α 858.1
914.2 930.2 946.2 962.2
1006.3 [M + Na]+
300 400 500 600 700 800 900 m/z
(D)
Y1 293.8
B2 370.8
444.8 467.8 * 484.8 504.8
544.8
636.1 *
720.9 *
Y2 776.0 792.9
Y2α 816.0
B3 853.0 871.0
Y3 908.1 943.0 948.0
Y3α 992.2
[M + Na]+ 1124.4
300 400 500 600 700 800 900 1000 m/z
(E)
Y1 293.8
386.8
467.8 *
B3 502.8 520.8
600.9 * 616.9 *
Y2 657.9 676.9 *
823.0 *
Y3 832.0
B4 867.1
Y2α 948.2
0,2X2α 990.3
Y4 1006.2
1062.3 1094.4
1138.4 [M + Na]+
300 400 500 600 700 800 900 1000 m/z
(F)
Y1 293.7 315.9
348.9
B2 370.7 390.9
446.0 *
461.9 467.7
B3 544.7 562.8
619.9 *
635.9 *
Y2 643.8 658.8 *
718.8 *
793.9 *
809.9 *
Y3 817.9
B4 864.9 895.0
Y2α 990.2
967.9 *
983.9 *
Y4 992.2
1166.4 [M + Na]+
300 400 500 600 700 800 900 1000 1100 m/z


(G)
293.8
370.8
386.8
467.9
484.8
544.8
562.9
642.9
657.9
718.9
736.9
816.1
832.0
849.0
865.0 *
909.1
990.2
1006.2
1032.3
1180.3 [M + Na]+
m/z
(H)
293.9
370.9
386.9
468.0
527.0
587.0
605.0
658.0
701.0
761.1
779.1*
832.1
865.0
874.5
907.1*
951.2
1032.3
1048.3
1068.7*
1102.3*
1222.4 [M + Na]+
m/z
(I)
293.8
348.8
391.0
445.8
505.8
526.8
586.8
604.8
649.9 *
657.8
700.9
760.9
778.9
832.0
891.0
907.0 *
951.1
1032.3
1048.3
1079.9 *
1222.4 [M + Na]+
m/z
(J)
293.8
328.7
364.9
386.7
445.9
467.7*
502.8
577.8
635.8
649.8 *
657.7
748.8
789.8
808.9 *
949.7
964.0
999.0
1080.2
1096.2
1122.3
1270.5 [M + Na]+
m/z


(K)
293.9
328.9
372.9
468.0
545.0
644.0
677.1
695.1
776.1
791.1*
851.2*
907.2
960.8*
982.8*
992.2
1027.2
1102.2
1122.4
1124.4
1158.8*
1202.4
1218.7
1259.9
1298.5 [M + Na]+
m/z
(L)
293.7
386.6
467.6
544.6
649.8
657.6
676.6
694.6
789.7
850.7*
947.8
980.8
1005.8
1040.8
1122.0
1138.0
1164.1
1312.2 [M + Na]+
m/z
(M)
293.7
448.0*
469.7*
544.7
577.8
615.7
635.8
657.7
676.7
718.7
736.8
753.3
789.8
850.8*
864.8*
924.9
963.9
997.0
1041.0
1122.2
1138.2
1312.4 [M + Na]+
m/z
(N)
293.7
348.9
364.9
386.6
403.9
447.9*
469.6
502.6
635.8*
649.7*
657.7
676.7
694.7
736.7
790.7*
831.8
850.8*
864.8*
883.6
925.5
963.9
1041.0
1122.3
1138.2
1164.4
1312.5 [M + Na]+
m/z


(O)
293.8
310.8
365.0
386.7
446.0
467.8
502.6
636.0*
649.8*
657.8
676.8
694.8
774.9
790.9*
831.9
850.9*
864.9
964.0
981.0
1006.1
1041.1
1122.3
1138.3
1164.4
1312.5 [M + Na]+
300
400
500
600
700
800
900
1000
1100
1200
m/z
(P)
293.7
370.6
544.6
657.5
718.6
736.6
831.6
832.6
892.6*
906.6*
1005.7
1038.8
1082.8
1164.0
1180.0
1206.1
1294.2
1312.4
1354.4 [M + Na]+
400
600
800
1000
1200
m/z
(Q)
293.8
364.9
386.7
418.9
445.9
635.8
649.9*
657.6
766.7
789.7
880.8
900.7
940.8*
983.5
1005.9
1079.7*
1131.1
1138.1
1212.3
1270.4
1402.4 [M + Na]+
400
600
800
1000
1200
m/z
(R)
293.8
364.8
386.7
445.8
467.7
635.8
657.7
676.7
789.7
831.7
850.8
868.8
963.9
1005.9*
1024.9*
1138.1
1171.1*
1215.2
1296.4
1312.4
1338.5
1486.5 [M + Na]+
400
600
800
1000
1200
1400
m/z


(S)
Y6 Y5 Y4 Y3 Y2 Y1
B1 B2 B3 B4 B5 B6
Y2α
Ume
Y1 293.8
364.8
386.7
445.8
467.7
Y2 657.7
635.8
B4 676.7
789.7 *
Y3 831.7
B5 850.8
868.8
Y4 963.9
1005.9 *
1024.9 *
Y5 1138.1
1171.1 *
B6 1215.2
Y2α 1296.4
Y6 1312.4
0,2X2α 1338.5
1486.5 [M + Na]+
400 600 800 1000 1200 1400 m/z
(T)
Y6 Y5 Y4 Y3 Y2 Y1
B1 B2 B3 B4 B5 B6
Y2α
Ume
Y1 293.7
B2 370.6
386.6
467.7
B3 544.6
641.7
Y2 657.6
718.6
734.6
789.7
815.7
Y3 832.5
B5 892.7
910.7
Y4 1005.9
1066.9 *
Y5 1180.1
1213.2*
B6 1257.8
Y2α 1338.5
Y6 1354.5
1380.5
1397.5 *
1528.6 [M + Na]+
400 600 800 1000 1200 1400 m/z

The PACE analysis of GH30 and GH10 products, and the MS analysis of GH30/GH3 were conducted over three occasions. Most of the results were reproducible; however, only representative results are shown. The hydrolysis with GH30 produced shorter oligosaccharides when the incubation was extended for longer time. However, the acetylation patterns on xylan described here were consistent across experiment replicates.

GH30/GH3 products on *gux1* and *gux2* xylan produced oligosaccharides that differ in the number of *O*-acetyl groups per backbone Xyl (degree of acetylation, DA). Grantham (2016) previously analysed the DA around the [Me]GlcA in the compatible domain and the incompatible domain by digesting *gux1* and *gux2* xylan with GH30 (or GH10), and analysed the masses of the oligosaccharides by MS. In the analysis performed here, GH3 was used after GH30, which may increase the DA of the products with the removal of unsubstituted Xyl. To compare the data obtained here with the previously reported, the five predominant GH30/GH3 products on *gux1* and *gux2* xylan with over four backbone Xyl are listed in Table 4.1. In the incompatible domain data, most of the incompatible fragments have DA over 0.5, whereas the compatible domain oligosaccharides have a DA of 0.5 or less. In the reported, *gux1* GH30 glucuronidated products had a DA over 0.5, while *gux2* GH30-product DA was approximately 0.5. The average DA obtained here are higher than the previously reported, because GH3 removes undecorated Xyl, raising the DA. However, the DA average of *gux1* is higher than *gux2,* which is consistent with that previously reported.

Table 4.1: Five highest abundance GH30/GH3 products on *gux1* and *gux2* with over four backbone residues.

domain	m/z	formula	DA
Compatible (*gux2*)	1270	$P_6Ac_3U^{me}$	0.50
	1312	$P_6Ac_4U^{me}$	0.67
	1534	$P_8Ac_3U^{me}$	0.38
	1576	$P_8Ac_4U^{me}$	0.50
	1618	$P_8Ac_5U^{me}$	0.63

domain	m/z	formula	DA
Incompatible (*gux1*)	1180	$P_5Ac_4U^{me}$	0.80
	1270	$P_6Ac_3U^{me}$	0.50
	1312	$P_6Ac_4U^{me}$	0.67
	1354	$P_6Ac_5U^{me}$	0.83
	1430	P_7Ac_4U	0.57

4.3 Discussion

In this chapter, to describe the acetylation in the incompatible and compatible domains, *gux1* and *gux2* xylan were hydrolysed and analysed by PACE and MS. The digestion products exhibited novel acetylation patterns, which are consistent and inconsistent with other patterns previously described.

4.3.1 Reproducibility and scope range

The activity of this specific GH30 xylanase does not tolerate a [Me]GlcA or Ac substitution at the subsite -1. Since acetylated Xyl contiguous to Xyl carrying a [Me]GlcA have been reported with low abundance (Chong et al., 2014), the oligosaccharides described in this chapter may represent only a fraction of the total xylan polymer. The description of GH10 hydrolysis products could cover overlapping fractions of the polymer with GH30 products. This is because, contrasting GH30 activity, GH10 does not require [Me]GlcA substitutions at a specific side of the hydrolysis cleavage point (see Chapter 1, section 1.5.2). Therefore, the use of GH10 would provide with further information about xylan acetylation patterns.

GH10 products were also analysed by mass-spectrometry (not shown), however, the results were found to be less reproducible. In addition, GH10 products seem to be more complex to analyse due to the high amount of isobaric structures, which could not be separated by HILIC. However, the mass-spectrometric approach described in this chapter could be further optimised for the analysis of GH10 products. A reverse-phase chromatography separation step could be considered as an alternative to HILIC for future xylan description approaches.

4.3.2 Pentose modification of the xylan backbone in *gux1* mutant may not originate from primary wall

The hydrolysis of *gux1* and *gux2* xylan with GH30/GH3 produced in both cases an oligosaccharide carrying a pentosylated GlcA (PUX). Mortimer et al. (2015) described the PUX structure to be characteristic of primary-cell-wall xylan. In agreement, Chong et al. (2015) also suggested that PUX originates from primary cell walls as this structure has higher abundance in stems of two mutants (*irx7* and *irx9*) that have a

reduction of xylan in secondary cell walls. Therefore, the pentosylated GlcA found in the GH30 products of *gux1* and *gux2* xylan may correspond to PUX and originate from primary-cell-wall xylan. Then, possibly *gux1* and *gux2* samples had a small proportion of primary cell wall. This could be tested by repeating the experiments in *gux1/3* and *gux2/3* mutants. This would exclude the possibility of GH30 hydrolysis glucuronidated products to originate from primary-cell-wall xylan. Interestingly, the PUX-modified species were acetylated (acetylation of primary-cell-wall xylan is discussed in Chapter 5).

However, the putative pentosyl sidechains found in this work on the secondary-cell-wall xylan backbone have not been described before in *Arabidopsis*. In Chapter 5, pentosyl decorations were detected on the callus xylan backbone, but none were detected on primary-cell-wall xylan from young stems. In this chapter, pentosyl sidechain decorations were found in *gux1* xylan, but none were detected in *gux2*, suggesting they may originate from the incompatible domain. One of GH30/GH3 hydrolysis products of *gux2* was also found in callus with low abundance; NXX^{aa}XU^{me,a}X (Figure 5.10A and D). The other pentosylated *gux2* product was not seen in callus; N^{a}X^{a}U^{me,a}X. Consecutive acetylated Xyl in the second structure also suggests it may have originated from the incompatible domain. It is a possibility that GH3 xylosidase may have not fully removed all backbone Xyl from the NRE on both oligosaccharides. However, this is unlikely since both *gux1* and *gux2* sample were digested in parallel, and no other GH30/GH3 products of *gux2* with unsubstituted pentoses at the NRE were found. To investigate the identity of these unsubstituted pentoses at the NRE, their sensitivity to GH62 arabinofuranosidase or GH3 xylosidase (see Chapter 1, Figure 1.30, page 46) would provide of the necessary information. If they are Ara*f*, the treatment with GH62 would produce a difference of 132 Da on the MS spectrum by the removal of this side chain. If they are non-hydrolysed Xyl at the NRE, these oligosaccharides could be further hydrolysed by GH3.

Pentosyl sidechains on xylan are found in other plant clades. α-L-Ara*f* substitutions on the xylan backbone are highly abundant in monocots and gymnosperms (Busse-Wicher et al., 2016b; Peña et al., 2016), and only a minor report has mentioned Ara*f* on xylan in *Arabidopsis* primary-cell-wall xylan (Tan et al., 2013). Xylan arabinosyltransferases belong to the GT61 family, and xylan xylosyltransferase activities have also recently been reported in this protein family (Zhong et al., 2018a).

Arabidopsis possesses genes that encode GT61 enzymes (Voiniciuc et al., 2015), and therefore, the pentosyl sidechains described here could be the product of a GT61 activity. If so, the identity of this pentosyl sidechain on the backbone could be either an arabinosyl or a xylosyl.

4.3.3 The xylan acetylation patterns are consistent with the 'domain compatibility'

The acetylation contexts around the patterned compatibly-spaced GlcA decorations on xylan by GUX1 (in *gux2* mutant), and the incompatibly-patterned GlcA substitutions by GUX2 (in *gux1* mutant) were different. The mass-spectrometric analysis of GH30/GH3 products of *gux2* xylan indicated that all acetylated Xyl near GUX1-mediated-[Me]GlcA are alternated by non-substituted Xyl, except for two minor products; $X^{a}X^{a}U^{[me],a}X$ (Figure 4.2E) and $X^{a}XXXU^{[me],a}$ (Figure 4.2G). In the compatible domain, the GlcA-modified Xyl and the acetylated Xyl were spaced by even number of backbone Xyl. The acetylation pattern observed extended from a [Me]GlcA-modified Xyl to up to six backbone Xyl towards the NRE. This acetylation pattern determined for the xylan compatible domain agrees with the main alternating pattern of acetylation previously reported (Figure 1.21; Busse-Wicher et al., 2014; Chong et al., 2014). Busse-Wicher et al. (2014) showed that xylan is acetylated mostly on alternated backbone residues. Their model is based on the predominant GH10 or GH30 products of WT or *gux1/2* xylan: $X^{a}XX^{a}X$. However, this is a non-glucuronidated oligosaccharide and its distribution in xylan molecules was not shown. Not long later, Chong et al. (2014) presented an alternating acetylation pattern model based on the main products of GH10 on WT xylan; $X^{a}X$ and $X^{aa}X$. They also described structures with mono- and di-acetylated Xyl two backbone residues towards the RE from a glucuronidated Xyl. They described the structures $U^{me,a}XX^{a}X$ and $U^{me,a}XX^{aa}X$, which complement the acetylation pattern of the compatible domain described here. Even distances between [Me]GlcA and Ac moieties agree with the xylan-cellulose interaction model presented by Busse-Wicher et al. (2014), in which all decorations can face away from the cellulose microfibrils.

The MS analysis of GH30/GH3 products of xylan regions that are decorated by GUX2 (in *gux1* mutant) revealed that the acetylation in the incompatible domain has patterns that are incompatible for the interaction with hydrophilic surfaces of cellulose. Several oligosaccharides showed consecutive acetylated Xyl, contrasting the pattern found in the compatible domain. These patterns involving consecutive acetylated Xyl, and Ac substitutions spaced by an odd number of backbone residues from other substitutions are consistent with the cellulose-xylan interaction model (Busse-Wicher et al., 2014), since they were all found in the region of xylan that is decorated by GUX2. GH30/GH3 products containing GlcA substitutions by GUX2 had less predominance for even number of backbone residues than of those decorated by GUX1. Consistent with this pattern, GH10 products on WT xylan described by Chong et al. (2014) exhibited consecutive modified Xyl; $XU^{me,a}X^aX$, $U^{me,a}X^aX^aX$ and $X^aU^{me,a}X^aX$, but the origin of these in the xylan molecules is unknown.

The incompatible acetylation pattern found in xylan regions decorated by GUX2 may be involved in the cross-linking of cellulose microfibrils. Since GUX2-mediated-[Me]GlcA substitutions have no preference for even or odd number of backbone residues of spacing, these regions of xylan are hypothesised to be incapable of interacting with hydrophilic surfaces of cellulose. This is because these interactions occur with xylan adopting a two-fold screw configuration (Simmons et al., 2016), and if substitutions are not at even distances from each other, they hinder the xylan docking onto this cellulose surface. This is why the regions of xylan decorated by GUX2 were defined as the incompatible domain (Bromley et al., 2013), however the pattern of acetylation in this region was not reported before. Here, two types of acetylation patterns were found in xylan decorated only by GUX2 (in *gux1* mutant); an ‘incompatible’ pattern that is consistent with the model of xylan-cellulose interaction (Busse-Wicher et al., 2014) and a ‘compatible’ pattern which is inconsistent with the model. In this incompatible pattern of acetylation, two or more mono- or di-substituted Xyl are arranged consecutively near the [Me]GlcA. This arrangement in a two-fold-screw configuration would display *O*-acetyl groups in two directions, which would hinder the interaction with cellulose. This incompatible acetylation pattern here described for xylan decorated by GUX2 might either directly prevent xylan from binding to the hydrophilic faces of cellulose, or influence GUX2 to make a pattern that is incompatible with binding. On the other hand, the compatible acetylation pattern

exhibited by xylan regions decorated by GUX2, may allow such interactions. This is because these regions exhibited Ac-substituted Xyl that were spaced by an even number of backbone residues of distance from each other, and from Xyl carrying [Me]GlcA, similarly to the acetylation pattern here described on the region of xylan decorated by GUX1.

The acetylation backgrounds around the [Me]GlcA substitutions in both xylan domains could elucidate GUX1 and GUX2 recognition sites on xylan associated with their specific activity. Possibly, GUX2 recognises consecutive mono-acetylated Xyl or regions with a high DA to transfer GlcA on to the backbone. In other way, GUX1 possibly recognises alternating acetylated Xyl with non-substituted Xyl. In agreement with the hypothesis that GUX1 recognises the alternating acetylation pattern of xylan, Grantham et al. (2017) proposed that GUX1 may recognise gaps on this pattern. Their analysis on PACE of the xylan acetyltransferase mutant *esk1*, which lacks the main alternating pattern of acetylation (Chapter 6; Grantham, 2016) shows no preference for even spacing between [Me]GlcA substitutions, indicating a dysregulation of GUX1. The acetylation pattern of xylan in *esk1* was investigated further in this thesis and are presented in Chapter 6.

Chapter 5: Primary-cell-wall xylan acetylation

5.1 Introduction

Xylan is found in primary cell walls of most plant species (Cosgrove, 2005), but the function of xylan in primary walls, especially in dicots, is unknown. Dicot primary-cell-wall xylan is particularly difficult to study because of its low abundance, compared to the large amount of xylan present in secondary walls. Primary-wall xylan chain length has been proposed to be shorter (M Busse-Wicher, unpublished) and its α-1,2-GlcA substitutions are repeatedly separated by a distance of six backbone Xyl (Mortimer et al., 2015). GlcA sidechains are also present and these can be further 4-*O*-methylated, or alternatively, and exclusively for primary-cell-wall xylan, α-1,2-arabinopyranosylated (PUX) (Mortimer et al., 2015; Peña et al., 2016). On the other hand, acetate groups on primary-cell-wall xylan have not been previously reported, except for the work of Chong et al. (2014) where the authors show one structure from *Arabidopsis* stem xylan in which a backbone Xyl is co-substituted with PUX and Ac. Since the PUX was subsequently shown to be primary wall specific, their data strongly suggests that primary-cell-wall xylan is acetylated, yet, further evidence is needed to get a better understanding of primary-cell-wall xylan acetylation.

[Me]GlcA substitutions on *Arabidopsis* primary-cell-wall xylan are dependent solely on the activity of GUX3 (Mortimer et al., 2015). In contrast, GUX1 and GUX2 show no contribution to the substitution of primary-wall xylan with PUX side-chains. Phylogeny of the GUX proteins indicate that GUX3 is more closely related to GUX1 than to GUX2 (Mortimer et al., 2010). The high similarity between GUX1 and GUX3 may help explain why the [Me]GlcA substitutions mediated by both enzymes are spaced by even number of Xyl. Possibly, GUX1 and GUX3 mechanism for substrate specificity and/or for patterning of [Me]GlcA onto the xylan is similar. However, GUX1 is influenced by acetylation context (Grantham et al., 2017), and the acetylation context in which GUX3 and GUX1 act *in vivo* may be different. The description of primary-cell-wall xylan acetylation may therefore help to understand the substrate specificity of GUX3.

In *Arabidopsis* stems, the proportion of xylan decorated by GUX3 (primary-cell-wall xylan) is higher in young over mature tissues (Mortimer et al., 2015). The enrichment of primary-wall xylan in younger tissues could have several reasons. Is it because

primary walls are synthesised first, and secondary cell walls later on by the same tissue? Alternatively, primary-wall xylans could be tissue specific; Xylem and phloem cells have thickened walls, but xylem walls are sclerenchymatous lignified non-living cells, while phloem cells are alive and permeable, despite their thickened wall (Dickison, 2000, chap. 2), and the phloem walls might be of a primary cell wall type. Basically, even though xylem and phloem are both vessels-type tissues, they have a different function, hence their cell wall constituents are different. Is the xylan content and structure also different between these tissues? Does the structure of primary-cell-wall xylan vary between tissues?

Callus is a system with only primary wall which is useful for xylan comparison. The *In vitro* culture of root *Arabidopsis* callus cells is a very interesting system to study primary cell walls since it is completely devoid of secondary-cell-wall structures. Callus cultures have been shown to have xylan but in very low amounts (Handford et al., 2003) and in agreement with this observation, primary-cell-wall-xylan-synthesis proteins are expressed and active in root-derived callus tissue (Nikolovski et al., 2012). During the progress of this work, Mortimer et al. (2015) showed that callus xylan hydrolysis with GH11 xylanase produces $U^{[me]}X_4$ and PUX_5, suggesting the presence of GlcA and PUX substitutions. However, the identity of these structures were not verified by mass-spectrometry. Other than this report, there are no publications describing callus xylan structure. Then several questions arise: Are callus xylan [Me]GlcA and PUX substitutions patterned? Do these substitutions depend on GUX3 activity? Is the GlcA methylated? Is this xylan acetylated? If it is, is there an acetylation pattern? Are there any other backbone modifications?

This chapter has three main aims oriented to study in greater detail primary-cell-wall xylan: First, to determine whether the different xylan [Me]GlcA patterns observed in primary growth are tissue specific. To do this, xylan from two early growth tissues from birch was investigated. One tissue was phloem in which the cells produce only primary walls, and the other was xylem in which the cells produce mostly secondary walls (Dickison, 2000). To study the substitution pattern of phloem and xylem xylan, GH30 products were compared by PACE and the structure of the two predominant phloem products were analysed by MS.

The second aim was to identify whether primary-cell-wall xylan from different sources vary in their structure. The structures found in birch for the first aim contributed to this purpose. In addition, callus xylan was also analysed. For this, callus primary-cell-wall xylan was analysed on PACE using GH30. Specific sidechains on GH30 products were detected by their sensitivity to glycosyl hydrolases with known specific activity: GH62 arabinofuranosidase and GH115 glucuronidase were used for detecting *Araf*, or GlcA with no further Ara*p* modifications, respectively. GH3 xylosidase was implemented to detect unsubstituted Xyl on the backbone NRE. GH30/GH3 products were then investigated further using mass-spectrometry, and the substitution patterns were determined by Laser-Induced Dissociation (LID) MS/MS. As a sub-aim, the dependency of callus xylan and its [Me]GlcA substitutions to active IRX10-L and GUX3 were also studied. To do this, xylan from *irx10-l* and *gux3* mutant calli were analysed by PACE using GH11 or GH30 xylanases.

The third aim was to describe the acetylation pattern of different types of primary-cell-wall xylan. To do this, the xylan was extracted with DMSO, which unlike alkali does not remove ester bonds from the polymer. *Arabidopsis gux1/2* stem and WT callus primary-wall xylan acetylations were analysed by three methods. First, Xyl mono-substituted with Ac were probed by testing the accessibility of GH11 to xylan when treated with CE4 acetylesterase. CE4 removes Ac groups from mono-acetylated Xyl which are not co-substituted with [Me]GlcA decorations (Biely et al., 2013; see Chapter 1, Figure 1.29 and Figure 1.30). Second, Ac on di-substituted Xyl were determined by analysing the migration shift on PACE of CE4/GH11 products when de-acetylating them with alkali. Third, to determine the acetylation patterns, the xylan was sequentially digested with GH30 and GH3, and the products were analysed by MS. The acetylated Xyl positions respect to the GlcA were determined by LID MS/MS.

During the progress of this work, analysis of the GlcA pattern was attempted using a similar approach to a previous report (Mortimer et al., 2015). Due to availability, GH30 from *Erwinia chrysanthemi* was used instead of *Bacteroides ovatus*. Interestingly, its products exhibited a singlet on PACE, which is different to what was previously described with the Bacteroides enzyme. The singlet corresponded to GlcA-modified xylooligosaccharides, without indication of PUX modifications. Therefore, a sub-aim of this chapter was to identify an available GH30 enzyme that can accommodate PUX

sidechain substitutions for hydrolysing xylan. Hydrolysis products of primary-wall xylan by GH30 enzymes from three different organisms were compared by PACE.

This work clarifies the Ac, [Me]GlcA and other substitution patterns present on xylan from primary cell walls, from two hardwood plants; *Arabidopsis* and birch.

5.2 Results

5.2.1 GH30 specificity for GlcA and PUX structures on xylan

To investigate the activity of different glucuronoxylanases, GH30 enzymes from three different sources were tested; *Dickeya dadantii* (formerly known as *Erwinia chrysanthemi*, EcGH30), *Bacteroides ovatus* (BoGH30, reported by St John et al. (2011)) and *Clostridium thermocellum* (CtGH30). To prepare the substrate, *gux1/2* plants were grown, newly grown stems (young stems) were collected, and the xylan was extracted with alkali. The choice of *gux1/2* means that only the primary-wall xylan was decorated with [Me]GlcA and was digested by these glucuronoxylanases. The extracted xylan was incubated with GH30, and then the enzymes were inactivated with heat. The products were labelled with ANTS and analysed by PACE (Figure 5.1).

Two predominant oligosaccharides were produced; the lower band corresponds to $U^{[me]}X_6$ which was produced by all three GH30. The upper band corresponds to PUX_6 and was only produced by BoGH30. Unexpectedly, CtH30 and EcGH30 digestion products did not show PUX_6 on PACE, suggesting these GH30s are incapable in these conditions of accommodating PUX sidechains for hydrolysing the xylan.

Additionally, BoGH30 digestion produced two less prominent bands (single asterisks). The identities of these oligosaccharides are unknown, but their detection suggests shorter distances of [Me]GlcA-substituted Xyl. The faint bands marked with two asterisks were possibly longer digestion products with a [Me]GlcA modification at position -2 from the RE, and a PUX sidechain on another backbone residue.

The different sensitivity exhibited by the GH30 activities to xylan PUX sidechains provided a tool for the detection of these arabinosylated GlcA. EcGH30 and BoGH30 enzymes were used later on in this chapter to study the sidechains in different types of primary-wall xylan.

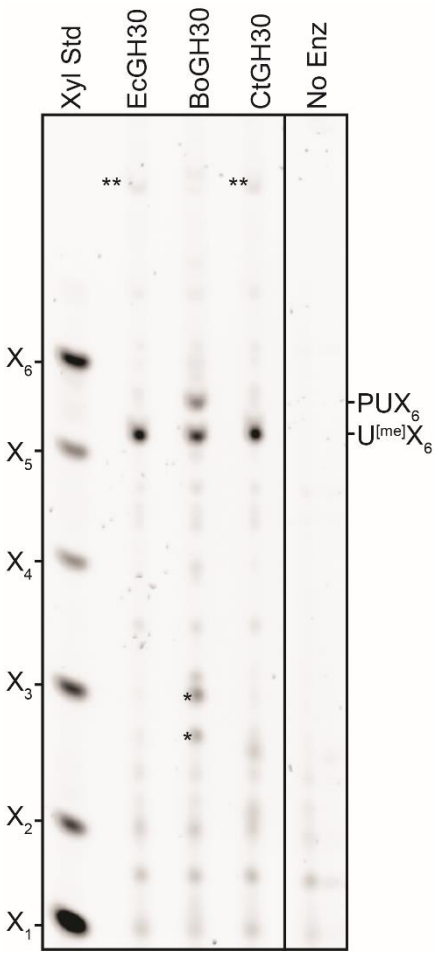


Figure 5.1: PACE analysis of GH30s from different species.
Xylan was extracted with alkali from *gux1/2* young stems, and digested with GH30 glucuronoxylanases from three different species: *Erwinia chrysanthemi* (EcGH30), *Bacteroides ovatus* (BoGH30) and *Clostridium thermocellum* (CtGH30). The products were labelled with ANTS and resolved on a 10% polyacrylamide gel. The undigested control (No Enz) was loaded on the same gel and middle lanes were removed. (*) and (**) are unidentified products.

5.2.2 Phloem cells produce uniquely primary-cell-wall-like xylan

The xylan structure differs between young and mature stems (Mortimer et al., 2015), therefore, perhaps xylan structure also varies between different tissues within an organ. To test this hypothesis, the [Me]GlcA substitution patterns of young phloem and xylem xylan were analysed and compared. In order to do this, young phloem and young xylem were cryo-dissected from birch (*Betula sp*). The material was kindly provided by Maija Tenkanen and Ykä Helariutta (University of Helsinki, Finland). *Arabidopsis* mature stem xylan was also added for comparison as a described [Me]GlcA pattern xylan control (Bromley et al., 2013). The xylan was extracted with alkali and digested with BoGH30. The digestion products were labelled with ANTS and analysed by PACE (see Figure 5.2).

Xylem xylan GH30 digestion produced oligosaccharides with a range of xylose residues, which co-migrated with *Arabidopsis* mature stem xylan products. Among the products with DP (degree of polymerisation) over eight, the ones with an even number of backbone residues were more abundant than the ones with an odd number of backbone residues. This suggests that the xylan [Me]GlcA pattern of birch xylem is similar to *Arabidopsis* secondary-cell-wall xylan, but the frequencies of the number of Xyl between [Me]GlcA decorations differ; the predominant species produced by xylem xylan was $U^{[me]}X_5$, while in *Arabidopsis* mature stem xylan was $U^{[me]}X_8$.

In contrast, phloem xylan produced only two predominant species. The product labelled with a single asterisk co-migrated with $U^{[me]}X_6$. The species marked with two asterisks did not co-migrate with any of *Arabidopsis* xylan products, but had a similar mobility to PUX_6 (Mortimer et al., 2015). This product was absent in technical replicates using EcGH30 instead of BoGH30 (not shown). The mobility of this oligosaccharide on PACE and the sensitivity to different GH30 glucuronoxylanases suggests that this phloem product was substituted with PUX sidechain, and to confirm its structure, it was analysed by mass-spectrometry.

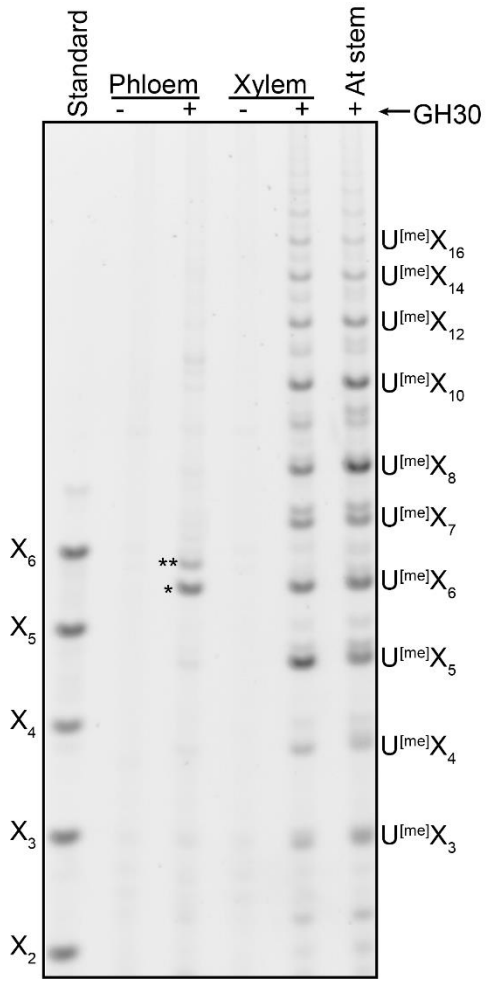


Figure 5.2: Tissue-specific GlcA of phloem and xylem from birch.
PACE analysis of birch phloem and xylem xylan using GH30. AIR was prepared from cryo-dissected material, and xylan was extracted with alkali and digested with BoGH30. The hydrolysis products were labelled with ANTS and resolved on a 20% polyacrylamide gel. (*) $U^{me}X_6$; (**) $PU^{me}X_6$. Representative of three biological replicates.

In order to analyse the birch-phloem GH30 products by mass-spectrometry, the oligosaccharides were derivatised with 2AA, and the spectra obtained are shown in Figure 5.3A. All *m/z* identified corresponded to sodiated adducts of the predicted masses ($[M+Na]^+$). The predominant species were *m/z* = 1144 ($U^{me}Pent_6$) and *m/z* = 1276 ($U^{me}Pent_7$). $U^{[me]}X_6$ observed on PACE correlated to *m/z* = 1144, as they appeared to be the most abundant species in the sample, confirming the identity of this oligosaccharide. The second strongest signal on the MS spectrum was *m/z* = 1276 ($U^{me}Pent_7$), but this species on PACE did not co-migrate with $U^{[me]}X_7$ from the *Arabidopsis* sample, implying that this oligosaccharide had a different isomeric structure.

To determine the structure of $U^{me}Pent_7$, a fragmentation analysis of this species was performed by MALDI-CID MS/MS. The spectrum obtained is shown in Figure 5.3C, and the two possible structures are presented in panel B. The spectrum corresponds to a mixture of both species shown, but $PU^{me}X_6$ was predominant over $U^{me}X_7$. The position of the MeGlcA substitution is on the second Xyl of the backbone from the RE in both structures, and this position was determined by the activity of GH30 (Chapter 1, Figure 1.1A). The ion *m/z* = 666.8 corresponds to an NRE elimination ion (E_5) and was diagnostic for $PU^{me}X_6$ since it can only be produced if the MeGlcA is on the fifth Xyl from the NRE. The ion *m/z* = 954.0 corresponds to a glycosidic bond cleavage ($Y_{2\alpha}$) and was diagnostic of the PU^{me} sidechain because it can only be produced if the MeGlcA carries a pentosyl residue. In the same way, the ion *m/z* = 682.9 (B_5) is an NRE elimination ion and was diagnostic for the presence of $U^{me}X_7$. *m/z* = 1086.0 ($Y_{2\alpha}$) is the fragmentation of the glycosidic bond for $U^{me}X_7$ diagnostic for the loss of a non-pentosylated MeGlcA. The abundance of both isomers could be inferred by three observations: First, the size of *m/z* = 954.0 ($Y_{2\alpha}$) compared to *m/z* = 1086.0 ($Y_{2\alpha}$) is indicative that $PU^{me}X_6$ was the predominant species. Second, the presence of E_5 and H_5 for $PU^{me}X_6$, plus the absence of E_6 and H_6 for $U^{me}X_7$, also suggest that $PU^{me}X_6$ was the predominant species. Third, in the PACE analysis of this sample (Figure 5.2), $U^{me}X_7$ was not visible. The fragmentation of the parent ion *m/z* = 1276 confirms the presence of the methylated PUX structure on xylan. No PUX structure carrying a methyl group has been previously reported in eudicots.

The PACE and MS analyses showed that phloem and xylem xylan differ in structure and pattern of [Me]GlcA substitutions. The $PU^{me}X_6$ structure was only found in the phloem samples and not in xylem. The distance between [Me]GlcA in phloem is therefore constant every six Xyl, while in xylem the spacing varies. The substitution pattern in phloem xylan and the methylated PUX structure have not been reported before. The structure and frequency of these substitutions on the birch phloem xylan are remarkably similar to the non-methylated PUX found in *Arabidopsis* primary-cell-wall xylan (Mortimer et al., 2015).

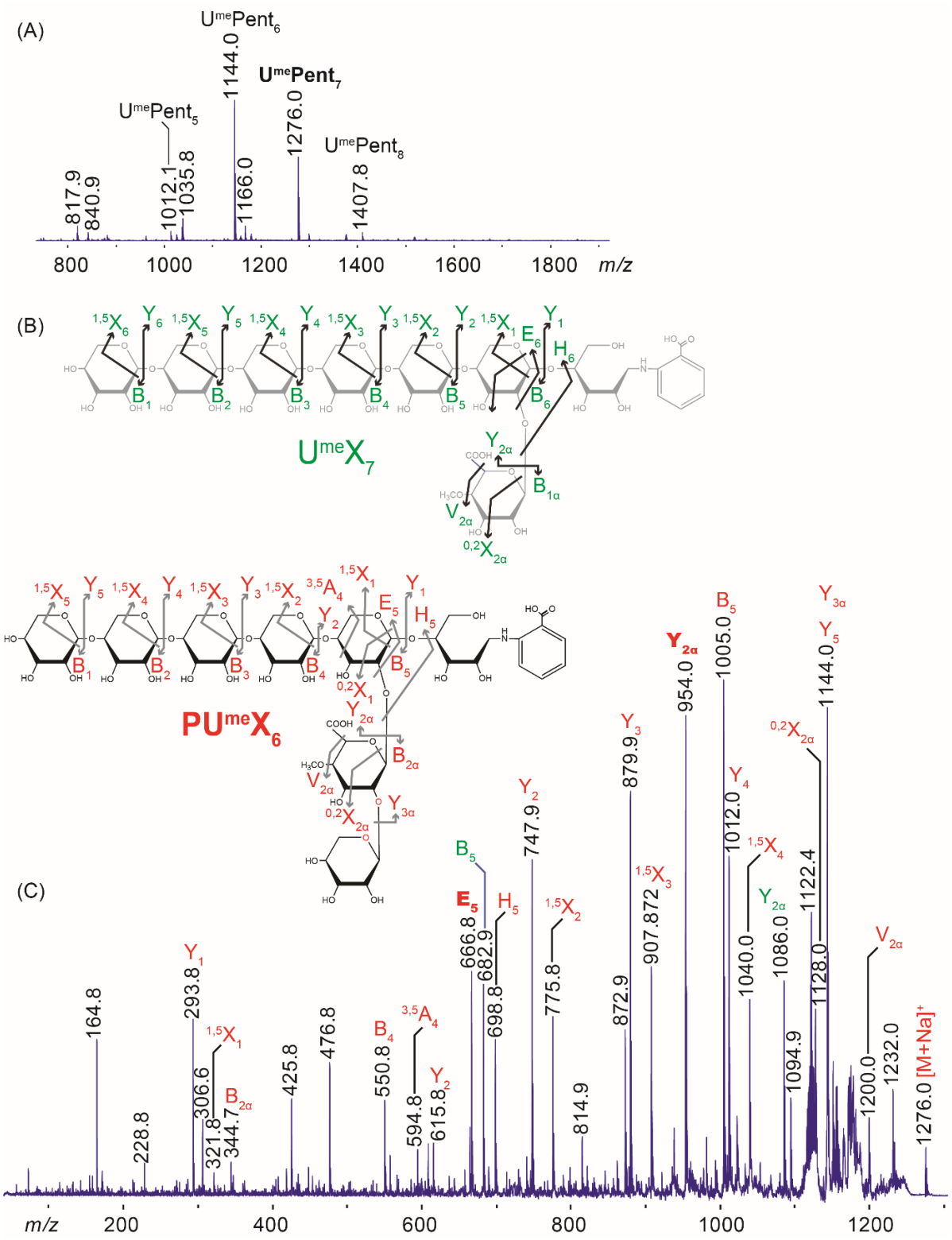


Figure 5.3: Mass-spectrometric analysis of BoGH30 products from phloem xylan. Mass-spectra of BoGH30 products on phloem xylan extracted with alkali and derivatised with 2AA. All oligosaccharides detected correspond to sodiated adducts $[M + Na]^+$. Spectrum shown in A was acquired in reflector mode. The spectrum of the molecular mass *m/z* = 1276 was acquired by CID-MS/MS and is shown in C. Two possible structural isomers are presented in B. Fragmentation ions are shown in green and red depending on the structure they were originated. The *m/z* calculation for H and E ions include the loss of 2 protons and a double bond formation.

5.2.3 *Arabidopsis* young stem primary-cell-wall xylan

5.2.3.1 Detection of Ac on xylan by PACE

Acetylation in primary-cell-wall xylan from *Arabidopsis* stems has been suggested (Chong et al., 2015; Mortimer et al., 2015), but the position of Ac modifications was unknown. To study primary-cell-wall xylan acetylation in stems, the detection of acetate groups were related to the remaining [Me]GlcA in *gux1/2* mutant xylan. In this mutant [Me]GlcA substitutions are dependent on the primary-cell-wall-xylan-specific glucuronidation by the GUX3 enzyme (Mortimer et al., 2015). The mutant *gux1/2/3* was used as a negative control for [Me]GlcA on primary-wall xylan. Newly grown stems were collected from both mutants as described by Mortimer et al. (2015). AIR was prepared and depectinated with ammonium oxalate. In order to determine whether primary-wall xylan has mono-acetylated Xyl, GH11 accessibility was compared when used solely or in combination with CE4 acetylesterase. Then, to determine whether acetate groups are on di-substituted Xyl (2,3-di-Ac-Xyl or (3-*O*-Ac)(2-*O*-[Me]GlcA)-Xyl), GH11 was removed by heat inactivation and purification with a Nanosep cartridge. Then the GH11/CE4 digestion products were treated with alkali. The oligosaccharides obtained were labelled with ANTS and analysed on PACE (Figure 5.4).

GH11 accessibility to *gux1/2* xylan increased when it was used in combination with CE4 in comparison to GH11 without CE4 (panel A). This is reflected by the increased production of X_1 and X_2, and the decrease in the higher DP products (asterisks) when CE4 was present. The production of bands A, E and F increased when CE4 was used, which implies that there are mono-acetylated Xyl on xylan which inhibit GH11 activity. Species A – F were not produced when digesting *gux1/2/3* xylan but were from *gux1/2* xylan, therefore, the mono-acetylated residues detected are associated with [Me]GlcA-modified oligosaccharides, originating in primary wall.

The identity of some oligosaccharides produced by GH11/CE4 were determined by their shift after treating with alkali. The identity of E is $XU^{[me],a}XX$, which after alkali treatment co-migrated with $XU^{[me]}XX$. After the alkali treatment, the structure D co-migrated with PUX_5, which suggests that structure D was acetylated. The mobility of other unidentified species (A, B, C and F) also shifted after alkali treatment (red

arrows). XX^{aa}XX was also found, which co-migrated with X_4 after the alkali treatment (blue arrow), however, this oligosaccharide could also originate from secondary-wall xylan, since it was also present in the digestion of *gux1/2/3* xylan.

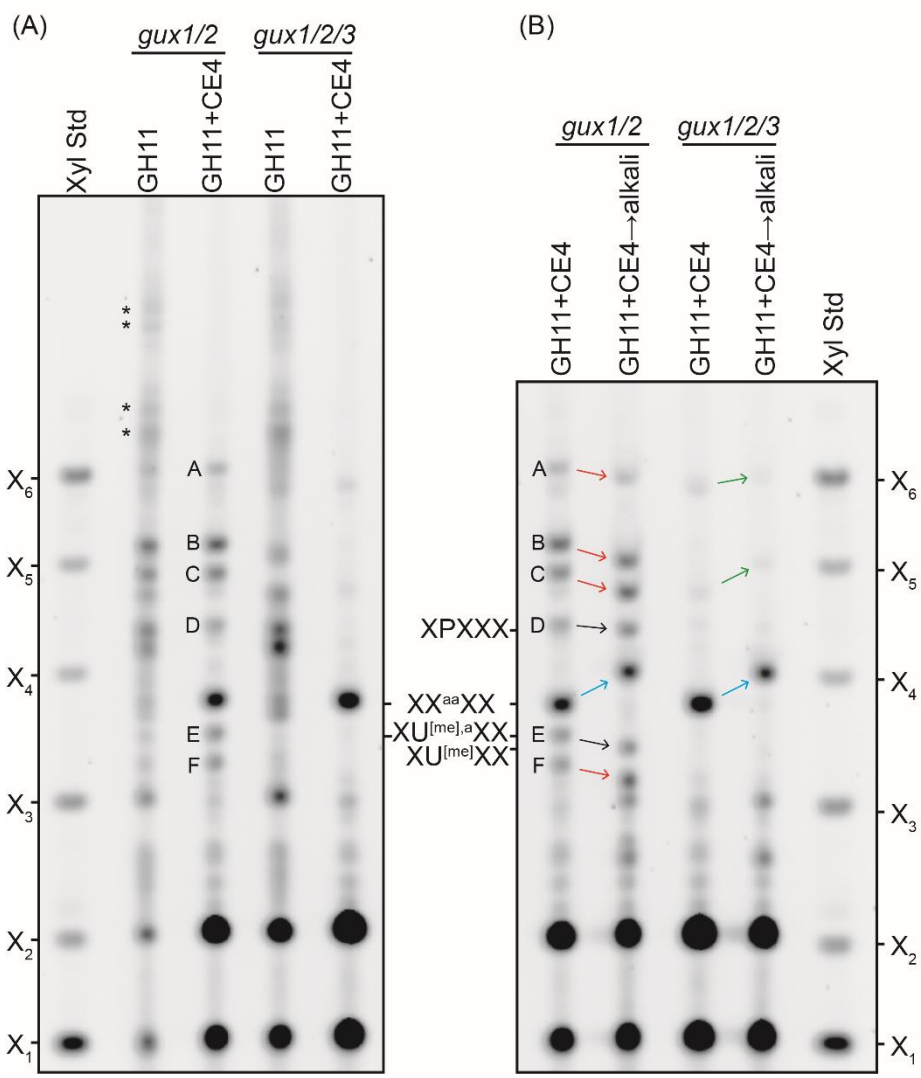


Figure 5.4: Detection of Ac modifications on *Arabidopsis* stem primary-wall xylan. *gux1/2* and *gux1/2/3* xylan was hydrolysed with GH11 in the presence or absence of CE4 (panel A). The GH11/CE4 products were then de-acetylated or not (+/-) with alkali (panel B). Then they were derivatised with ANTS, and resolved on 10% PACE gels. GH11/CE4 products unique for *gux1/2* were labelled A – F. Arrows in (B) show shifting bands when treated with alkali. Green arrows show the shift of unknown species without GlcA, and Blue arrow show the shift of XX^{aa}XX to X_4. The GlcA modified species that were identified after alkali treatment are shown with a black arrow. Red arrows show the shift of unidentified species modified with GlcA.

In summary, these experiments showed that more than one form of Ac are present in stem primary-wall xylan. Some Xyl are possibly mono-substituted with Ac, and other acetylated Xyl may carry a second substitution such as an additional Ac or [Me]GlcA. However, Ac substitutions on the xylan sidechains are also possible. To determine the arrangement of the acetate groups in primary-wall xylan a mass-spectrometric approach was performed, which is shown next.

5.2.3.2 Determination of the xylan acetylation pattern in stem primary walls

The acetylation pattern of *Arabidopsis* stem primary-cell-wall xylan was determined by relating the Ac positions to the [Me]GlcA substitutions mediated by GUX3. To do this, *gux1/2* newly grown stems were collected as described by Mortimer et al. (2015). AIR was prepared and depectinated, and the xylan was extracted with DMSO. The xylan was then sequentially hydrolysed with BoGH30 glucuronoxylanase and GH3 xylosidase, with a heat incubation step between digestions to inactivate the xylanase. GH30 products have a 1,2-linked [Me]GlcA on the second Xyl from the RE, and GH3 xylosidase removes all unsubstituted backbone Xyl from the NRE, reducing the number of isomers for the following MS analysis (see Chapter 1, Figure 1.31B). To remove the acetylated neutral oligosaccharides that might arise by partial digestion of secondary cell wall acetyl xylan (Busse-Wicher et al., 2014), the acidic oligosaccharides were purified by solid phase extraction and analysed by mass-spectrometry. The spectrum of the digestion products is shown in Figure 5.5 A. The molecular structure of the predominant species (in green) were sequenced by Laser-Induced Dissociation (LID), and the spectra and structures are presented in panels B – F.

To understand the variety in the substitution patterns on stem primary-wall xylan, the structure of GH30/GH3 products is next discussed individually.

The most abundant species was $Pent_5Ac_3U$ ($X^{aa}XP^{a}X$, *m/z* = 1003), and its structure is shown in panel C. The glycosidic bond cleavages $Y_{2\alpha}$ (*m/z* = 695) and $Y_{3\alpha}$ (*m/z* = 871) were diagnostic for the presence of Ara*p* on the GlcA which was not methylated (PUX structure). Y_3 (*m/z* = 655) diagnosed the position of the di-acetylated Xyl at the NRE, and Y_2 (*m/z* = 522) determined that the second Xyl from the NRE was unsubstituted. The difference between B_2 and B_3 showed that the Xyl carrying PUX had an Ac modification. This species had four backbone Xyl, and the substituted Xyl were alternated with unsubstituted Xyl.

The fragmentation analysis of the species *m/z* = 885 ($X^{aa}XU^{me,a}X$, Panel B) revealed that its structure was relatively similar to *m/z* = 1003 ($X^{aa}XP^{a}X$, above). The ion $Y_{2\alpha}$ (*m/z* = 695), however determined that this species had an *O*-methyl group on the GlcA but not an Ara*p* modification. The difference between B_1 and B_2 (*m/z* = 371 and 735), and also the presence of Y_2 (*m/z* = 536) determined the presence of acetylation on

the glucuronidated Xyl. Y_3 (*m/z* = 669) showed that the third Xyl from the RE was unsubstituted, and that the Xyl on the NRE was di-acetylated. Therefore this structure had alternating substituted backbone residues.

The second most abundant species was *m/z* = 1309 ($X^{a}XX^{aa}XP^{a}X$, panel F). The glycosidic bond cleavage $Y_{2\alpha}$ (*m/z* = 1001) was diagnostic of a PUX sidechain on this structure. Y_5 (*m/z* = 1135) denoted the position of a mono-acetylated Xyl at the NRE. Then Y_2 and Y_3 (*m/z* = 655 and 787) showed that the Xyl carrying a GlcA was followed by an unsubstituted Xyl towards the NRE. Y_4 (*m/z* = 1003) determined that the fourth Xyl from the RE was di-acetylated, and followed by an unsubstituted Xyl towards the NRE. This species had six backbone residues which were alternately substituted.

The species *m/z* = 1191 ($X^{a}XX^{aa}XU^{[me],a}X$, panel D) and 1267 ($X^{a}XX^{a}XP^{a}X$ panel E) were less abundant. In panel D, the series Y_5, Y_4, Y_3 and Y_2 (*m/z* = 1017, 885, 669 and 537) diagnosed the contiguous residues $X^{a}XX^{aa}X$~, while B_5 and B_4 (*m/z* = 1041 and 677) showed the RE side of the molecule; ~$U^{me,a}X$. In panel E, B_5 and B_4 (*m/z* = 1093 and 635) denoted ~$P^{a}X$ on the RE, and the series Y_5, Y_4, Y_3 and Y_2 (*m/z* = 1093, 961, 787 and 655) characterised the sequence $X^{a}XX^{a}X$~ on the NRE. The modifications of GlcA sidechains in structures D and E were determined by the respective $Y_{2\alpha}$ (*m/z* = 1001 and 959). These less abundant species had both six backbone residues and alternated substituted Xyl in their backbones.

In summary, all structures determined had an even number of backbone residues, either four or six. The spacing between [Me]GlcA substitutions in *gux1/2* young stem xylan was of six Xyl (Figure 5.1 and section 5.2.2, above; Mortimer et al., 2015), however two structures had only four backbone Xyl (B and C). This indicates that the activity of GH3 may have removed two backbone Xyl from the GH30 products, suggesting that these oligosaccharides had two additional unsubstituted Xyl at the NRE. In contrast, the structures with six backbone Xyl were not substrates for GH3 (xylanase) because of their mono-acetylated Xyl at the NRE. All structures analysed had a pattern of alternating substituted Xyl, including Ac and [Me]GlcA. Also, all (methyl)glucuronidated Xyl were also acetylated.

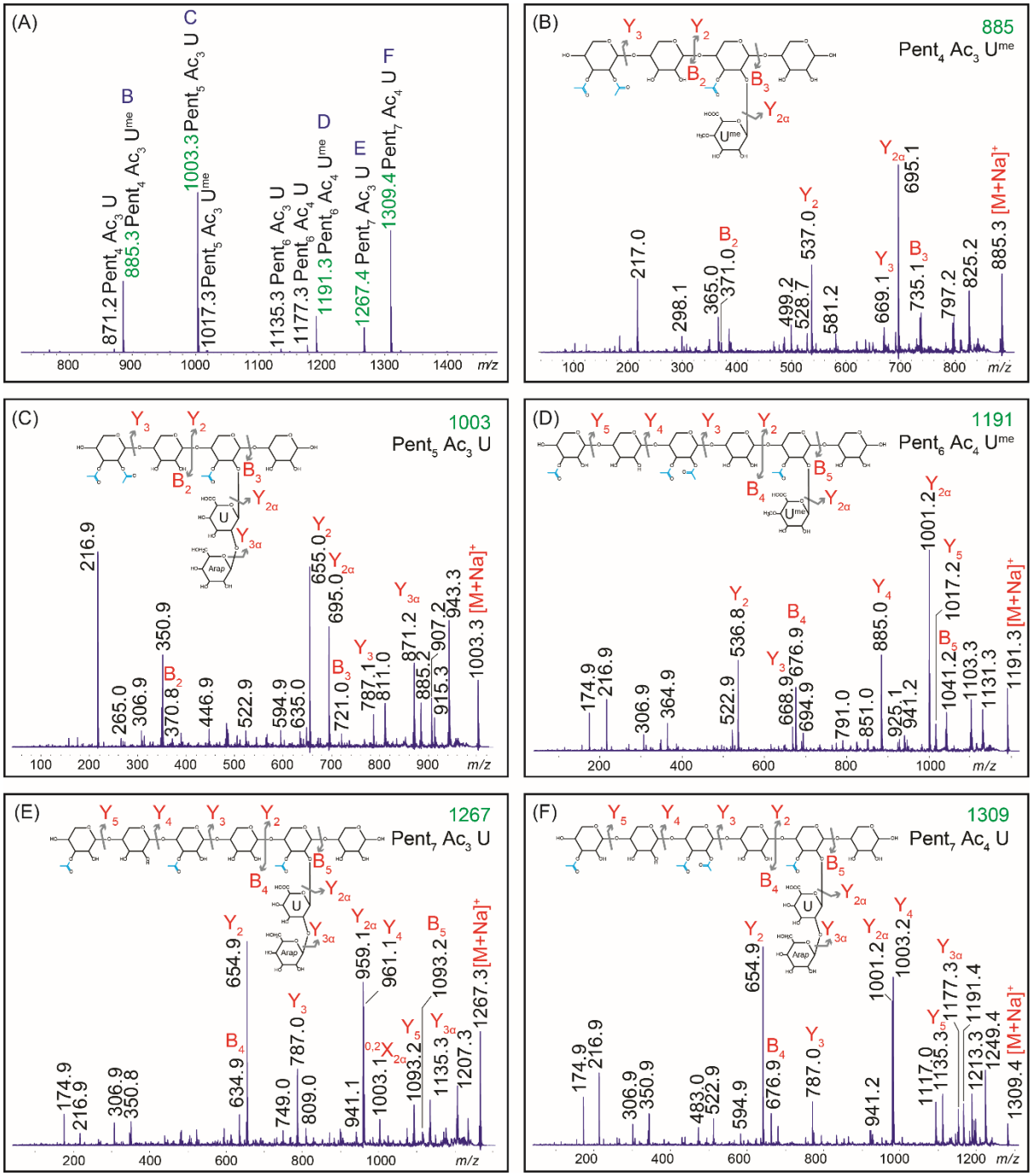


Figure 5.5: MS spectra of young stem acetylated xylan digested with GH30 and GH3. *Arabidopsis gux1/2* young stem xylan was extracted with DMSO, and then sequentially digested with BoGH30 and GH3. The products were cleaned with a C18 column and analysed by MALDI-ToF MS (A). The major products obtained (in green) were further analysed by MALDI-LID MS/MS (B-F). The positions of the Ac on mono-acetylated Xyl were not determined, but for simplicity they were drawn O-3-linked.

5.2.4 Callus xylan

Callus is an interesting system to study primary cell walls since it is completely devoid of secondary walls. However, callus xylan sidechains, pattern and functions are not fully understood. The presence of PUX structure identified in xylan from callus suggests that this xylan has some similarity to stem primary-wall xylan (Mortimer et al., 2015), however it does not necessarily mean they are identical. Here, a description of callus xylan was carried out in two steps. The first step consisted of the identification of glycosyl side chains and determination of the pattern in the backbone (section 5.2.4.1). The second step was to identify and describe any callus xylan acetylation pattern (section 5.2.4.2). To support the view that callus primary-cell-wall xylan biosynthesis is similar to the stem primary-wall xylan, the dependency of callus xylan backbone and GlcA substitutions to the activities of IRX10-L and GUX3 was investigated (section 5.2.4.3).

5.2.4.1 *Arabidopsis* callus xylan has pentosyl substitutions on the GlcA and on the backbone

To analyse the GlcA substitution pattern of callus xylan, fresh subcultures were grown for three weeks. Then AIR was prepared, and the xylan was extracted with alkali/$NaBH_4$ at 37 °C overnight (see section 2.4.3.2 in Chapter 2). The borohydride was removed by washing twice with 70% ethanol. This extraction method was found to be more efficient than using only alkali for 1 h at room temperature. Then the extract was incubated with EcGH30 or BoGH30 (which have different sensitivity to PUX sidechains on xylan; section 5.2.1). The digestion products were labelled with ANTS and resolved on PACE (Figure 5.6). EcGH30 produced similar signals to BoGH30 products, indicating that these [Me]GlcA substitutions were not further modified with Ara*p*. To determine the identity of the GH30 products, they were then tested for sensitivity to specific glycosyl hydrolases.

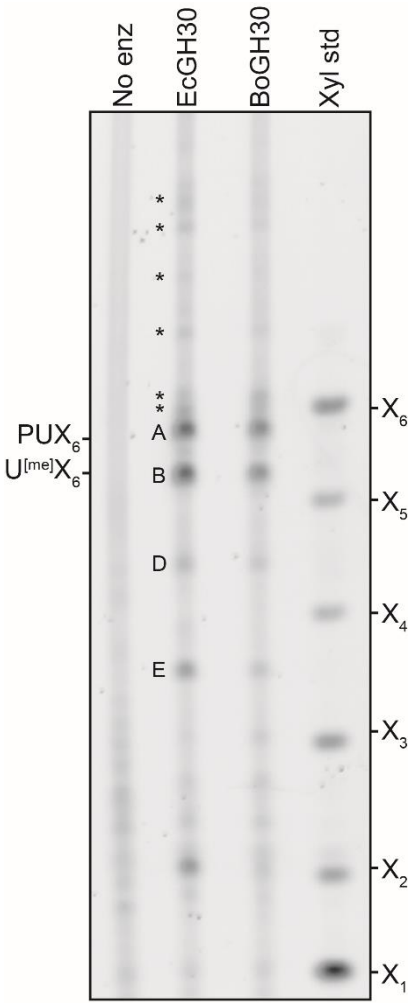


Figure 5.6: PUX structure detection in xylan from callus.
Xylan from callus was extracted with alkali/$NaBH_4$ and then digested with EcGH30 or BoGH30 glucuronoxylanases. The products were labelled with ANTS and resolved on a 10% polyacrylamide gel. The mobility of species A and B are different relative to the Xyl standard markers to PUX_6 and $U^{[me]}X_6$ (shown in Figure 5.1). Species highlighted with letters were produced by both enzymes. Species marked with asterisk were only visible in EcGH30.

To identify the oligosaccharides produced by EcGH30, other carbohydrate active enzymes were used on the products of EcGH30. To get a cleaner xylan extract without traces of borohydride, callus AIR was depectinated with ammonium oxalate, and the xylan was extracted with DMSO. The extracted xylan was treated with alkali which breaks all ester links, removing all possible Ac. Then the xylan was hydrolysed with EcGH30. The enzyme was inactivated with heat, and the products were then incubated with either GH62 arabinofuranosidase, GH115 glucuronidase or GH3 xylosidase. The products were then labelled with ANTS and resolved on PACE (Figure 5.7).

EcGH30 hydrolysis generated six bands on PACE (marked A – F), and the sensitivity to GH62, GH3 and GH115 is explained next.

Bands A, B, D, E and F were sensitive to GH115, indicating they had a [Me]GlcA substitution. Band C was partially affected by the treatment with GH115, which suggests that band C may correspond to two species that co-migrated on PACE, and one of them possessed a [Me]GlcA.

Species A and C were both insensitive to GH3, which suggests that both species were likely to have a modification on xylan that makes the NRE inaccessible to GH3. In addition, both bands shifted after the treatment with GH62, indicating that both species had Ara*f* on the backbone. Therefore, it is possible that A and C had an Ara*f* substitution at the NRE Xyl. This would imply that some Ara*f* substitutions on the original polymer backbone are two or three Xyl towards the RE from a [Me]GlcA.

Species B and D showed no sensitivity to GH62, indicating that they had no mono-arabinofuranosylated backbone Xyl.

The species A and C after GH62 incubation, co-migrated with B and C, suggesting that the structures of A and B, and C and D may have differed only by the presence of Ara*f* substitutions.

Band E co-migrates with UX_4, it showed insensitivity to GH62, and shifts after GH3 incubation, therefore, it is likely that the identity of band E corresponds to $XXU^{[me]}X$. The production of band E, therefore indicated that some [Me]GlcA substitutions on callus xylan are spaced by four backbone Xyl.

Band F was insensitive to GH62, and after GH3 its abundance was increased, which suggests that the identity of this oligosaccharide was $U^{[me]}X$. The migration of $U^{[me]}X$ previously identified is consistent with that exhibited by the band F (Dr Theodora Tryfona, personal communication). The production of band F by GH30 indicates that some [Me]GlcA substitutions on callus xylan are distanced by two backbone Xyl. The increased production of band F by GH3 indicates that some GH30 products with higher DP only had a single sidechain that corresponds to [Me]GlcA.

The detection of Ara*f* on callus xylan therefore suggests that primary-wall xylan from callus and stem may be different. To describe the callus xylan structure more fully, it

was necessary to perform additional experiments. Since the bands' intensity was low, it would be useful to replicate the analysis. The identity of some of the structures shown here were studied by MS, in combination with the analysis of acetylation on callus xylan later in this chapter (section 5.2.4.2).

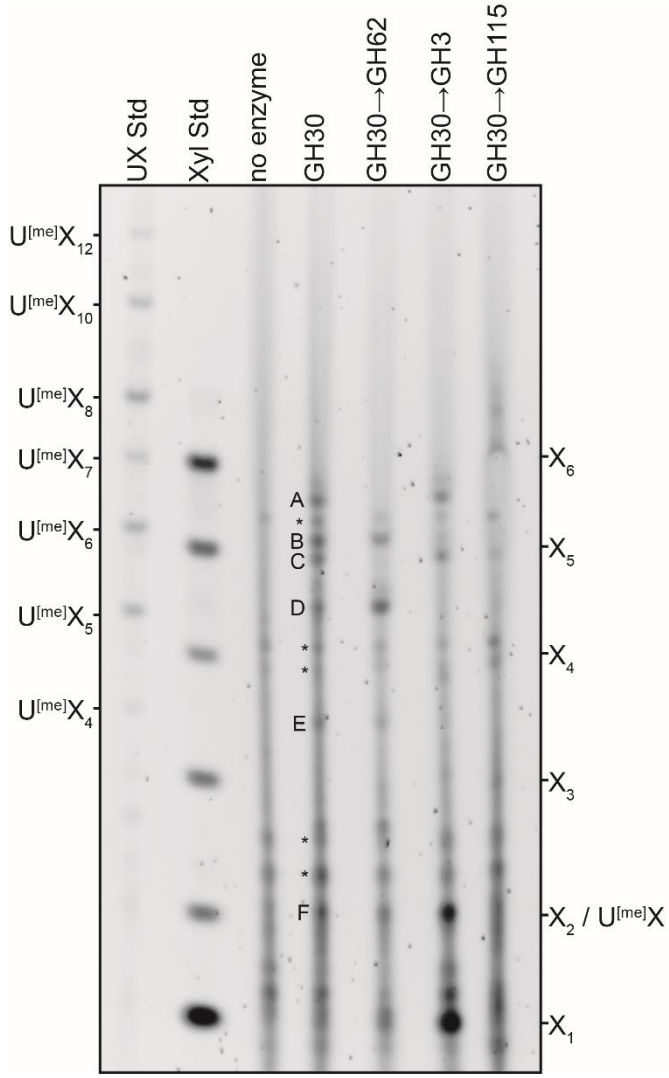


Figure 5.7: Further digestions on GH30 products from callus de-acetylated xylan. Callus xylan was extracted with DMSO and then treated with alkali. The polymer was then digested with EcGH30 and the products were incubated with GH62 arabinosidase, GH3 xylosidase or GH115 glucuronidase. The digestion products were analysed in a 10% polyacrylamide gel. GH30 products were marked with A – E, and contaminant bands were labelled with asterisks. One biological replicate.

The PACE analysis of GH30 products on alkali-extracted xylan suggests that Ara*f* and GlcA substitutions on xylan in some cases may be spaced by one or two Xyl. If this is true, GH11 should not be able to hydrolyse between the two substitutions. In order to confirm this, DMSO extracted xylan was treated with alkali and incubated with GH11 xylanase. The products were desalted with a HyperSep column and derivatised with

procainamide to increase their mass and to label the RE to aid interpretation of the MS/MS spectra. Then the oligosaccharides were mixed with DHB and analysed by MS (Figure 5.8 panel A). The molecular structure of the predominant species (in green) were sequenced by LID, and the spectra and structures are presented in panels B – F. The structures shown were determined following the Y ion series for each MS/MS spectrum.

Some GH11 products with more than one substitution denote the existence of neighbouring sidechains. The most abundant GH11 product was *m/z* = 1350 ($XU^{me}XNXX$, where N is a pentosyl substitution linked to the backbone) and its structure is shown in panel C. Its structure had seven backbone residues with a MeGlcA and a pentosyl substitution on the backbone. The spectrum showed that the MeGlcA and Pent decorations on this structure were spaced by two backbone Xyl. The structure of *m/z* = 1468 (XPXNXX, panel E) had a PUX sidechain and a Pent spaced by two backbone Xyl. In a similar manner, the structure of *m/z* = 1482 (panel F) had a methylated PUX sidechain at two backbone Xyl of distance from a pentosyl sidechain. All three structures (C, E and F) had the Pent towards the RE from the [Me]GlcA, denoting a preferred orientation between the two. However, *m/z* = 1336 (not shown) seemed to be a mixture of XUXNXX and XNXUXX, but its MS/MS spectrum signal was too weak for interpretation. The structure of *m/z* = 1408 ($XU^{me}XU^{me}XX$, panel D) had two MeGlcA sidechains neighbouring with each other, spaced by two backbone Xyl. The production of oligosaccharides with more than one sidechain by GH11 confirmed that some callus xylan substitutions are close to each other. Also, the spacing of two backbone Xyl between substitutions suggests they may have a cellulose-binding compatible backbone substitution pattern.

Some sidechains modifications on callus xylan were isolated from other glycosyl substitutions. The production of *m/z* = 954 ($XU^{me}XX$, panel B) by GH11 revealed that some [Me]GlcA substitutions were isolated. This structure had four backbone Xyl, and therefore, could only be produced by GH11 if a MeGlcA substitution on the xylan substrate backbone is spaced by four or more backbone Xyl from another glycosyl sidechain. The structure of *m/z* = 896 was possibly XNXX, which is produced by GH11 from isolated backbone pentosyl modifications. However, this structure was not determined, and its abundance was very low. Possibly, glycosyl sidechains may not

be very frequently isolated from other sidechains on callus xylan, especially pentosyl modifications.

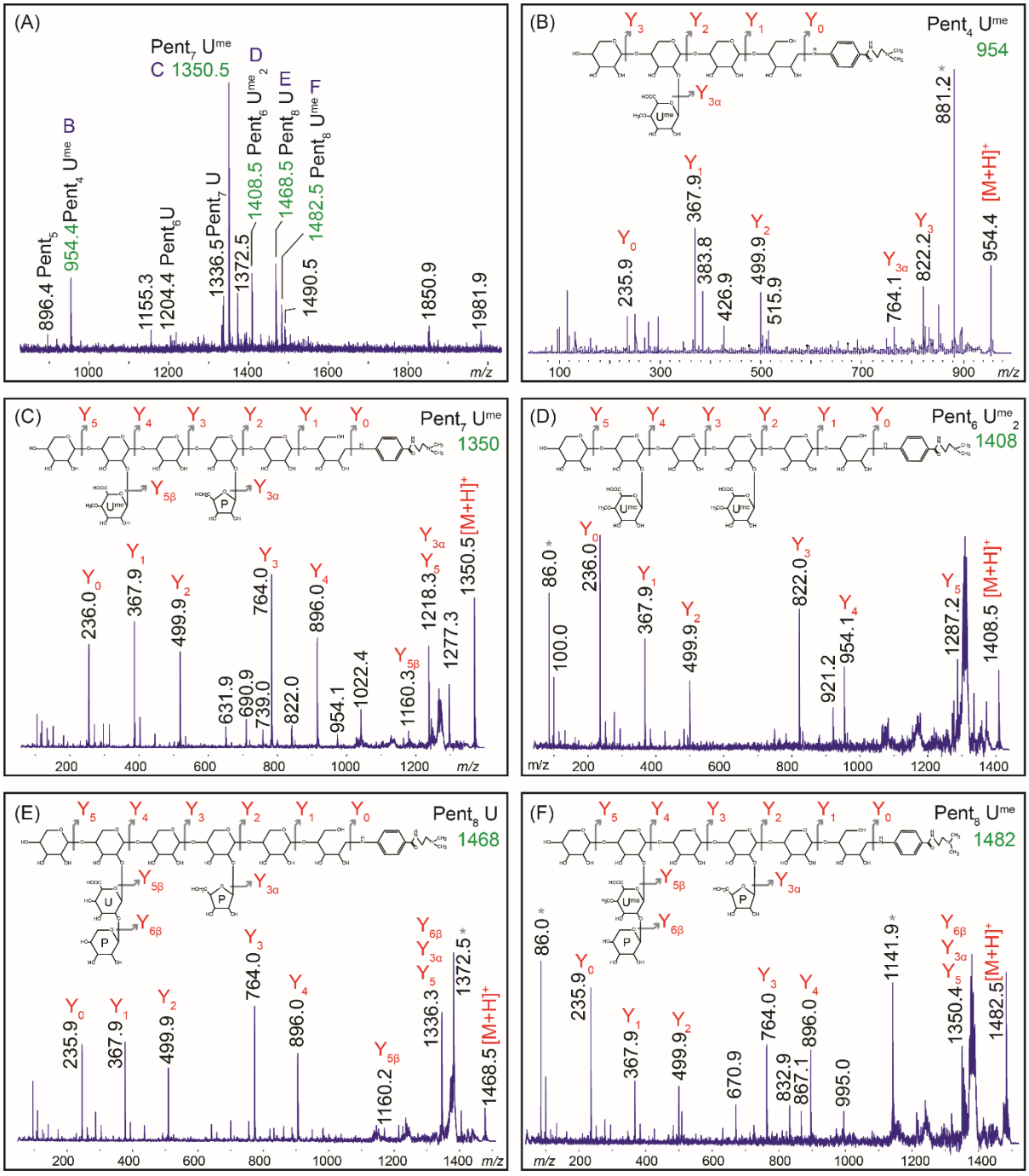


Figure 5.8: Mass-spectra of callus xylan treated with alkali and digested with GH11. Xylan was extracted with DMSO and then treated with alkali. The products of the GH11 digestion were desalted using a HyperSep column, labelled with procainamide and mixed with DHB. In panel A, GH11 products mass-spectrum is shown, and adducts are protonated. Labels B – F correspond to species in the other respective panels. Most abundant structures are shown in green, and LID MS/MS spectra and structures are shown in panels B to F. Grey asterisks show unknown peaks. The identities of the pentosyl substitutions and their linkages to the backbone Xyl were not determined, but were drawn as 1,2-linked pentofuranosyl sidechains for simplicity.

5.2.4.2 Determination of the acetylation pattern of callus xylan

The detection and analysis of callus xylan acetylation were carried out similarly to the analysis of primary-wall xylan from stem in section 5.2.3. First, the sensitivity of callus xylan to CE4 acetylesterase and alkali were used to determine the presence of Ac. GH11, and GH11/CE4 products were compared on PACE to detect mono-acetylated Xyl, and the shift of GH11/CE4 products when treated with alkali was used to detect Ac on di-substituted Xyl. Second, to determine the acetylation pattern of xylan, GH30/GH3 products were sequenced by MALDI-LID MS/MS.

For the analysis on PACE, a fresh callus subculture was grown for three weeks, AIR was prepared and depectinated with ammonium oxalate. Then the xylan was extracted with DMSO and digested solely with GH11, or in combination with CE4. GH11 was removed from the incubations by heat-inactivation, and by purifying the products with a Nanosep column. Then the GH11/CE4 products were treated where indicated with alkali. The oligosaccharides were labelled with ANTS and resolved on PACE (Figure 5.9).

GH11 accessibility to callus xylan increased when used in combination with CE4. This is identified by three observations: First, the increased production of X_1 and X_2 which indicates the increased hydrolysis of backbone residues. Second, the production of species D and E was increased. Third, the species marked with a black asterisk was reduced when treated with CE4. CE4 specificity for Ac on mono-substituted Xyl (Biely et al., 2013), and the increased GH11 accessibility to xylan when exposed to CE4 implies that callus xylan is acetylated and also that it possesses mono-acetylated Xyl on the backbone.

The mobility on PACE of some GH11/CE4 products shifted when treating them with alkali. The blue arrow shows the shift of $XX^{aa}XX$ to X_4, the black arrow shows the shift of $XU^{[me],a}XX$ (E) to $XU^{[me]}XX$. The species marked with a green asterisk has also shifted, but the position of this oligosaccharide prior to the alkali treatment was not determined. These shifts indicated the presence of Ac on di-substituted Xyl on the xylan, both di-acetylated and also Ac on [Me]GlcA-substituted Xyl.

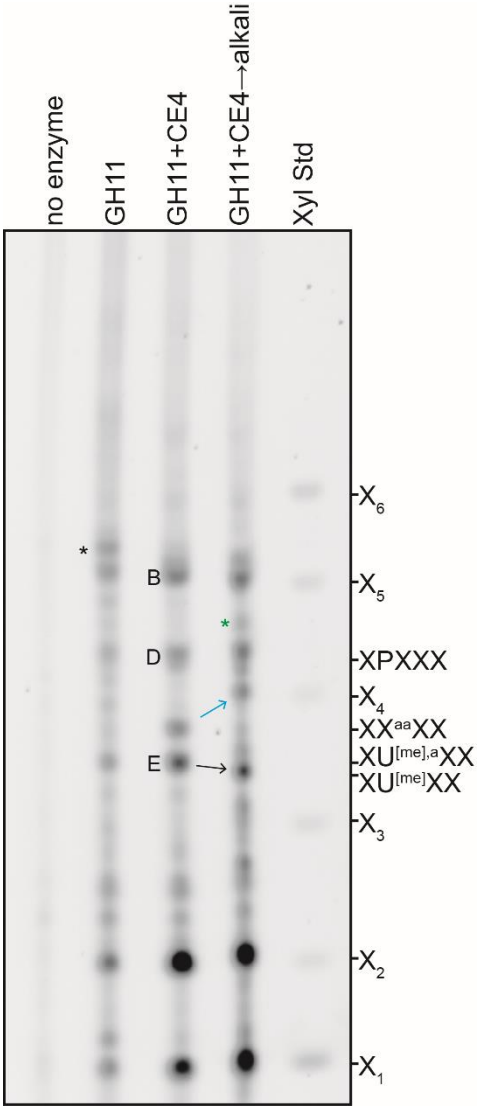


Figure 5.9 Ac detection on callus xylan by PACE.
Callus xylan was extracted with DMSO. The xylan was then incubated with GH11 or co-incubated with GH11 and CE4. The GH11/CE4 products were purified with a Nanosep column and then treated (or not) with alkali. The oligosaccharides were labelled with ANTS and resolved on a 10% PACE gel. The band labels B, D and E correlate with the ones used in Figure 5.4. Asterisks point to unknown species. Blue and black arrows show the shifts of $XX^{aa}XX$ and $XU^{[me],a}XX$ when treating them with alkali.

In order to identify the acetylation pattern of callus xylan, GH30/GH3 products were sequenced by mass-spectrometry following the procedure outlined above for primary-wall xylan from stems (section 5.2.3.2). The spectra are shown in Figure 5.10. In panel A, the spectrum of the sample that was used as input for HILIC is shown, and the structures identified by LID-MS/MS are presented in panels B – G. The structures shown were determined by following the Y and B ion series (Domon and Costello, 1988) for each MS/MS spectrum.

To consider the variety in the substitution patterns on callus xylan, the structure of GH30/GH3 products is next discussed individually.

The predominant species was *m/z* = 1360 ($NXX^{a}XU^{me,a}X$, $Pent_7Ac_2U^{me}$), and its structure is shown in panel C. The glycosidic bond cleavage $Y_{2\alpha}$ (*m/z* = 1170) was diagnostic of the MeGlcA substitution. The difference between Y_1 and Y_2 (*m/z* = 294 and 658) indicates that the Xyl carrying a MeGlcA was also acetylated. Then, Y_3 and Y_4 (*m/z* = 790 and 964) denote a mono-acetylated Xyl spaced by two backbone residues from the MeGlcA modification. The mass differences between Y_4 and Y_5 (*m/z* = 964 and 1096), between Y_5 and Y_6 (*m/z* = 1228) and between Y_6 and the total molecular mass, is 132. This indicates that this structure had three successive non-acetylated pentoses at the NRE. If GH3 hydrolysis was total, the insensitivity of the last pentose to GH3 xylosidase implies that this was a pentosyl sidechain on the NRE backbone Xyl. Therefore, this oligosaccharide likely had six backbone residues, and the substituted backbone Xyl were alternated with unsubstituted Xyl.

The structures D – G also had a difference of 132 between their respective Y_4, Y_5, Y_6 and their total mass. As they were all resistant to GH3, therefore, they also likely had a pentosylated NRE Xyl, followed by an undecorated Xyl.

The structure of *m/z* = 1402 ($NXX^{aa}XU^{me,a}X$, panel D) was similar to *m/z* = 1360 (above), but the acetylated Xyl carried an additional Ac. The ions Y_1, Y_2 and Y_3 (*m/z* = 294, 658 and 790) had the same values of the ones in structure C, indicating that structure D had similar substitutions on its RE trisaccharide (~$XU^{me,a}X$). The difference between Y_3 and Y_4 (*m/z* = 1006) is 216, indicating that the fourth Xyl from the RE carried two Ac.

The second most abundant species was *m/z* = 1492 (NXN^{a}XU^{me,a}X), which is shown in panel F. The ion $Y_{2\alpha}$ (*m/z* = 1170) indicates that the MeGlcA in this structure was not further modified, and the difference between Y_1 and Y_2 (*m/z* = 294 and 658) shows that the Xyl carrying the MeGlcA was acetylated. The difference between Y_3 and Y_4 (*m/z* = 790 and 1096), and the difference between B_3 and B_4 (*m/z* = 419 and 857) are both diagnostic of a di-substituted Xyl that carried a Pent and an Ac.

m/z = 1478 (NXX^{a}XP^{a}X, panel E) had a similar number of pentosyl residues to *m/z* = 1492 (above), however the position of one of the pentosyl sidechains was different. In this structure, $Y_{2\alpha}$ with *m/z* = 1170 (and not *m/z* = 1310) indicates the further modification of the GlcA with a pentose (PUX) since this glycosidic bond breakage can only be produced by the presence of PUX. Then the difference between Y_1 and Y_2 (*m/z* = 294 and 776) indicates that the GlcA-substituted Xyl and the Pent-substituted Xyl were both acetylated. The difference Y_2 and Y_3 (*m/z* = 776 and 908) is 132, therefore, the following backbone Xyl to P^{a}, was undecorated. Then the difference between Y_3 and Y_4 (*m/z* = 1082) is 174, indicating that the next Xyl was mono-acetylated, at two backbone residues from the GlcA modification. The structure F had a pentosyl sidechain on an acetylated Xyl, and in structure E the Ac position does not vary, but the pentosyl was on the GlcA.

The structure *m/z* = 1610 (NXN^{a}XP^{a}X, panel G) was the largest structure identified, and its backbone had alternating substituted Xyl with non-substituted Xyl. The $Y_{2\alpha}$ ion (*m/z* = 1302) denotes the further modification of the GlcA with a pentosyl residue. Then Y_1, Y_2 and Y_3 values were similar to structure E, therefore the RE of structure G was also ~XP^{a}X. The difference between Y_3 and Y_4 (*m/z* = 908 and 1214) indicates that the fourth backbone Xyl from the RE was co-substituted with an Ac and a Pent.

The smallest structure identified was *m/z* = 1006 (X^{aa}XU^{me,a}X, panel B), with only four backbone Xyl. Its $Y_{2\alpha}$ ion (*m/z* = 816) indicates that the MeGlcA substitution on the backbone was not further modified. The difference between Y_1 and Y_2 (*m/z* = 294 and 658) show that the GlcA was on the second Xyl from the RE, and that it was on an acetylated Xyl. Y_3 shows that the contiguous Xyl was unsubstituted, and therefore, the NRE Xyl had two Ac. It was not determined whether the precursor of the structure B was generated by GH30 alone, or whether GH3 reduced the backbone Xyl to four.

All structures analysed were consistently acetylated and/or pentosylated on alternating backbone residues. In all structures, the second and fourth backbone Xyl from the NRE were acetylated. The second Xyl from the RE therefore, was mono-acetylated, and co-substituted with a [Me]GlcA. The fourth backbone Xyl was either mono-acetylated, di-acetylated, or co-substituted with an Ac and a pentosyl residue. The pentosylated Xyl on the NRE were spaced by two Xyl from the nearest substituted Xyl, and no Ac were found on pentosylated Xyl at the NRE.

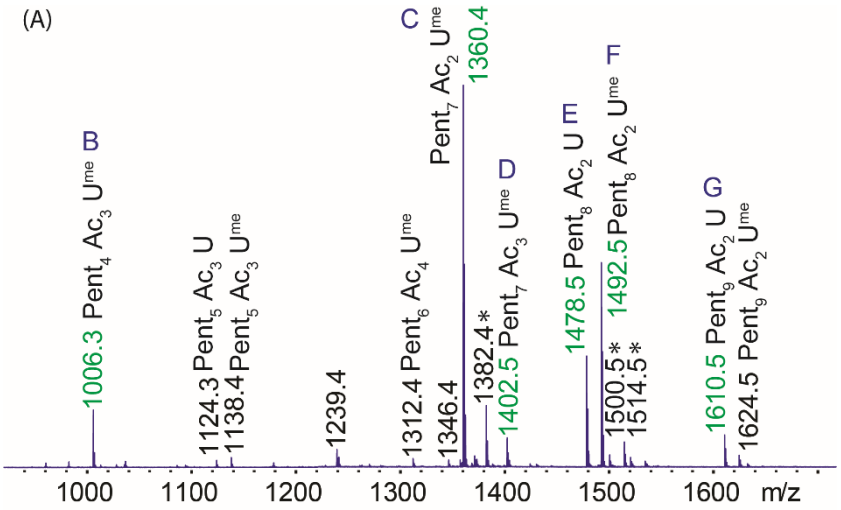


Figure 5.10: Structures containing Ac in callus xylan.
Arabidopsis callus xylan was extracted with DMSO, and then sequentially digested with BoGH30 and GH3. The hydrolysis products were purified with a C18 column, labelled with 2AA and analysed by MALDI-ToF MS (A). The linkages of the Xyl substitutions were not determined, but for simplicity they were drawn as 1,2-linked pentosyl modifications, and 3-linked-Ac on Xyl. The black asterisks in panel A point to peaks corresponding to doubly sodiated adducts from m/z = 1060.4, 1478.5 and 1492.5 respectively. Blue asterisks in panel E point to peaks obtained for contaminant isomers with lower abundance. The structure shown in panel E is the predominant structure found for m/z = 1478. (continued).

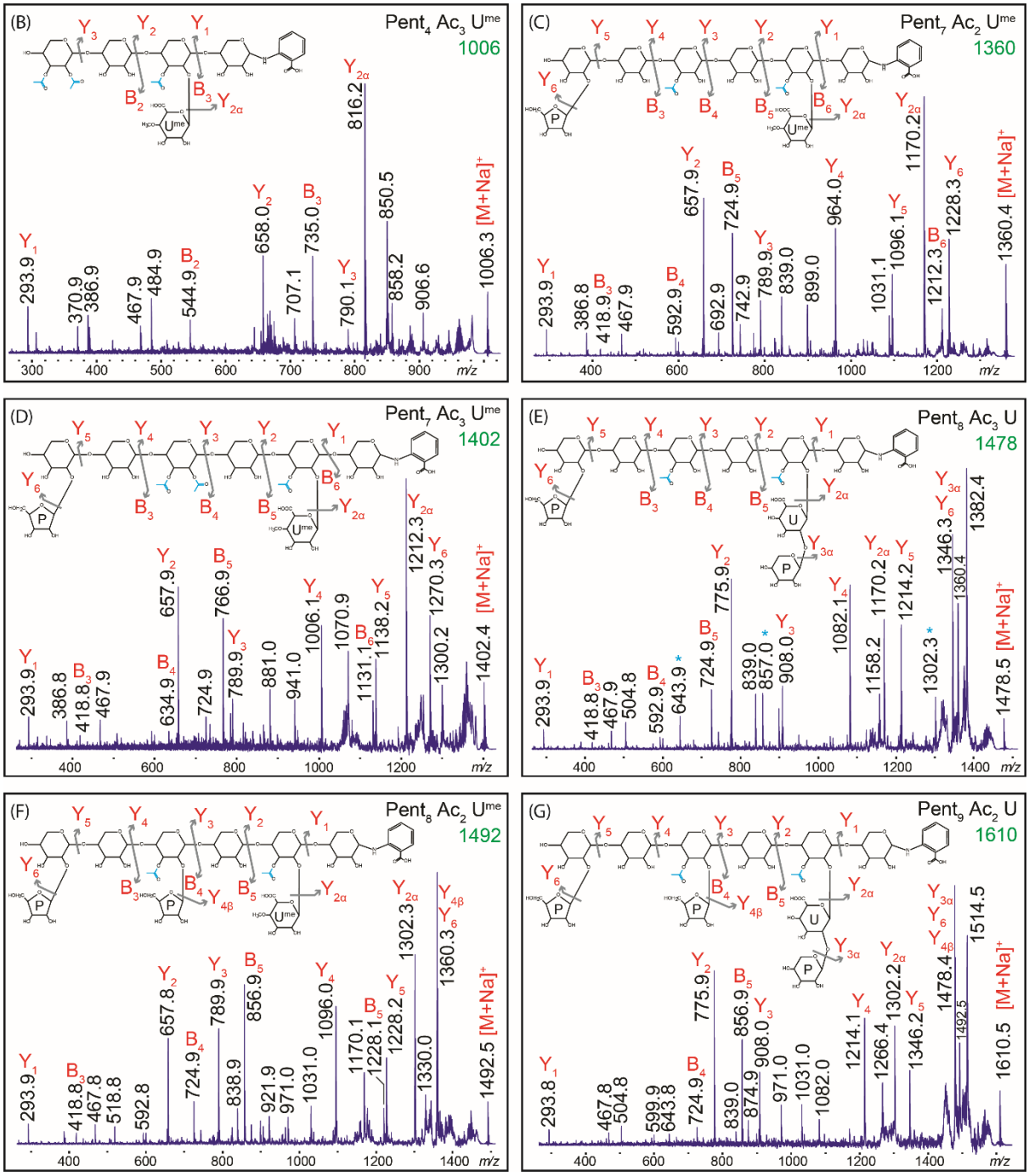


Figure 5.10. (continued).

5.2.4.3 GlcA containing oligosaccharides are GUX3 dependent

To study the dependency of callus xylan and its GlcA substitutions to IRX10-L and GUX3 which were previously shown to be necessary for stem primary-cell-wall xylan synthesis (Mortimer et al., 2015), the xylan in respective mutants was analysed by PACE. For this, 3-week-old calli subcultures were grown by Xiaolan Yu, and AIR was prepared. Equal amounts of AIR were quantified and xylan was extracted with

alkali/Na_2BH_4 at 37 °C. The borohydride was then removed with two 70% ethanol washes. Then the xylan was incubated with EcGH30 or GH11. The digestion products were labelled with ANTS and resolved on PACE (Figure 5.11A and B). To aid interpretation, the densitometric values of the indicated bands in panel A were quantified using the software GeneTools and used for estimating the amount of Xyl (panel C) and calculating relative amounts of each product (panel D).

GH11 xylanase produced a complex pattern of oligosaccharides as described above (section 5.2.4.1). The amount of xylan released by GH11 digestion of *irx10-l* AIR was decreased compared to WT (panels A and C). In WT and *gux3*, X_1 and X_2 were the predominant products, but in *irx10-l* they were greatly decreased. The presence of GH11 products on *irx10-l*, however, indicates that despite the inactivation of IRX10-L, this mutant has xylan xylosyltransferase activity in callus.

The relative abundance of some GH11 products was affected in *gux3* and *irx10-l* (panel D). A decreased production of GH11 products when GUX3 is inactivated was diagnostic of them carrying a [Me]GlcA substitution. Therefore, as E and G were barely undetectable in *gux3*, it is likely that they possess a GUX3 GlcA substitution. In addition, E and G were also nearly undetectable in *irx10-l* sample, thus, they probably arise from xylan made by IRX10-L. The mobility of species F suggests it is $XU^{[me]}XX$, and since its production was partially reduced in both mutants. This implies that some callus xylan is synthesised by a XSC with IRX10, and substituted by other glucuronyltransferases than GUX3.

The predominant GH30 products seemed dependent on GUX3 activity (panel B). WT callus xylan hydrolysed with EcGH30 produced the species marked with the letters H and I. BoGH30 also produced these two bands when used in similar conditions (section 5.2.1). H and I products were undetectable on PACE on *gux3* and *irx10-l* xylan hydrolyses, suggesting they are dependent on the activity of GUX3, and consequently dependent also of IRX10-L. To understand why GH30 did not produce oligosaccharides that carry GlcA substitutions and GH11 did, it may be useful to test the reproducibility of the hydrolyses, and also to determine the oligosaccharides structure by MS.

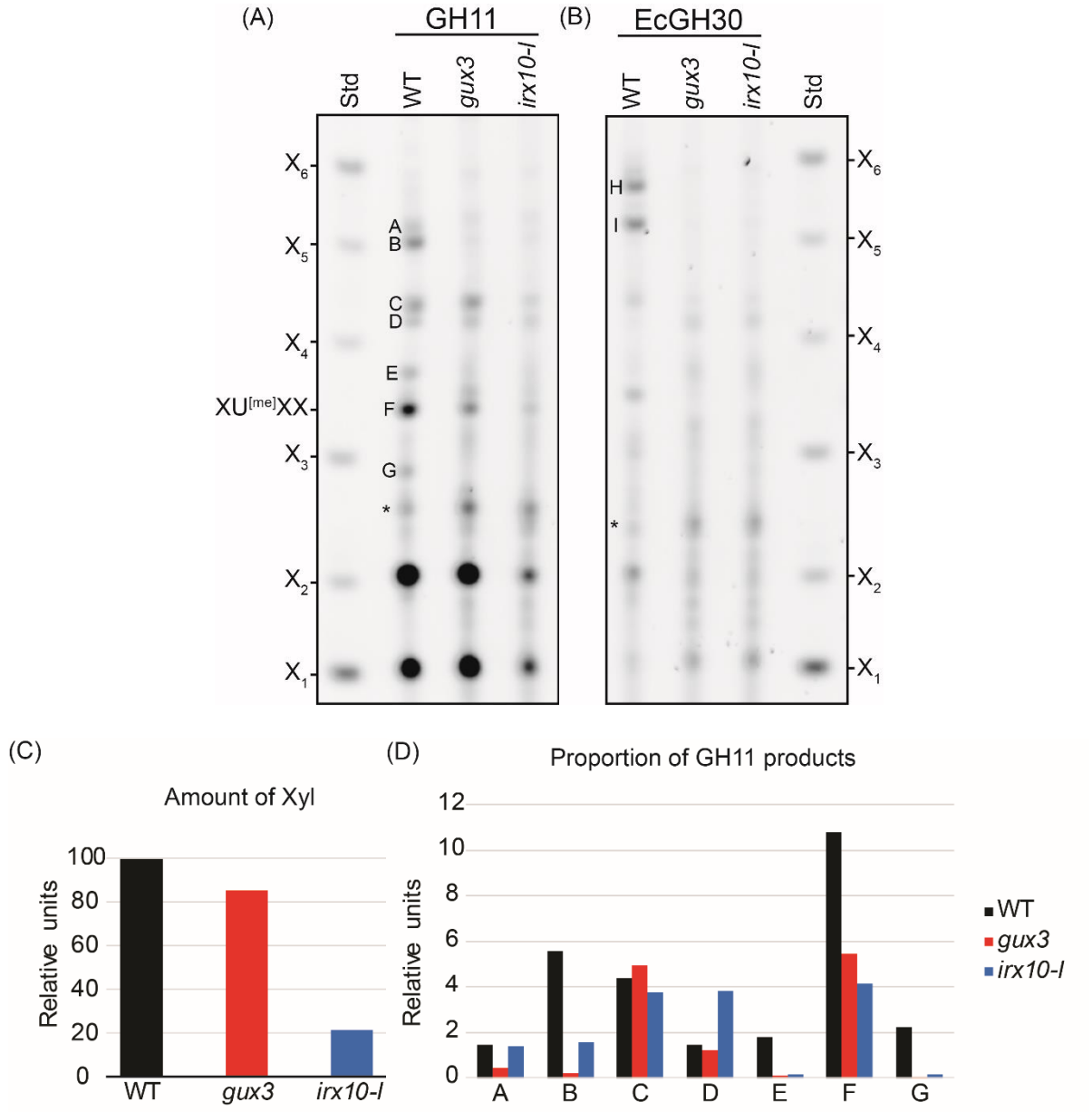


Figure 5.11: Detection of xylan and [Me]GlcA substitutions in callus mutants by PACE. *irx10-l* and *gux3* mutants and WT calli were grown for three weeks and AIR was prepared. Equal amounts of AIR were used for extracting xylan was extracted with alkali/Na_2BH_4 at 37 °C. GH11 (panel A) and EcGH30 (panel B) products were labelled with ANTS and resolved on 10% polyacrylamide gels. To aid interpretation figures in panels C and D were constructed with a densitometric quantification of species X_1, X_2 and species A – G from this gel. Panel C shows the relative amounts of Xyl released by GH11. The relative amount of Xyl was calculated by adding the densitometric value of X_1 plus twice the value of X_2 plus four times the value of F (since its likely to be $XU^{[me]}XX$), and the totals were normalised to the WT. Panel D shows the densitometric values of GH11 products A – G, normalised to the amount of Xyl. The species H and I in panel B are species marked with A and B in Figure 5.6, respectively. Asterisk: possible contaminant. Representative of one replicate.

5.3 Discussion

The three aims of this chapter focused on a better understanding of the structure and diversity of primary-cell-wall xylan. First, to determine whether xylan structures and substitution patterns are tissue-specific within an organ, birch phloem and xylem xylan was analysed. These two vascular tissues exhibited contrasting [Me]GlcA substitution patterns on xylan, which relate to the predominance of primary and secondary walls in each, respectively. The second aim was to identify whether the xylan structure and its glycosidic substitution patterns vary between primary cell walls. Ara*f* substitutions were only identified on callus xylan, which indicates a substantial difference to the primary-wall xylan from stem. The third aim was to describe the acetylation pattern of primary-wall xylan. As callus and stem primary-wall xylans from *Arabidopsis* were different, they were both analysed and compared. The acetylation pattern found on both xylans are analogous and are discussed below in more detail. Together, the results discussed here illustrate different levels of substitution patterns, and show that there is more than one type of primary-cell-wall xylan.

From this work, three main conclusions can be drawn. First, the substitution patterns found in primary-cell-wall xylan suggest they can interact with cellulose hydrophilic surfaces in a two-fold screw configuration. Second, the arrangement of Ac and pentosyl modifications on different types of primary-cell-wall xylan suggest they may have similar roles. Third, the substitution pattern of phloem xylan relates to the xylan decorated by GUX3 in *Arabidopsis* stems. Each of these conclusions is discussed in the following sections.

5.3.1 Primary-cell-wall xylan is compatible for interaction with cellulose

Perhaps the most intriguing result obtained in this work is that all primary-cell-wall xylans have the potential to form interactions with cellulose, following the model proposed by Busse-Wicher et al. (2014) (see section 1.4.4.3). All substituted backbone Xyl, either with Ac, [Me]GlcA, PUX, PU^{me}X, Pent, or the combinations of these were found alternated with undecorated Xyl. This is consistent with all primary-cell-wall xylan having the ability to interact with cellulose in a two-fold screw conformation, in a manner similar to xylan decorated with GUX1. The strongest evidence for this can be

observed from the mass-spectrometric analyses of *gux1/2* stem and WT callus xylan when hydrolysed with GH30/GH3 (glucuronoxylanase/xylosidase).

First, the structure of all *gux1/2* xylan GH30/GH3 products showed, without exceptions, an even number of backbone Xyl of distance between acetylated Xyl and (methyl)glucuronidated Xyl. Acetylated Xyl were always at two backbone residues from each other. Also, the (methyl)glucuronidated Xyl were acetylated, and therefore, the acetylation pattern is 'in frame' with the [Me]GlcA substitutions.

Second, callus xylan also showed only an even number of backbone Xyl between substitutions. The GH30/GH3 products analysed by MS showed constantly a decorated Xyl every two backbone residues on all callus structures determined. Acetylated Xyl were at two backbone Xyl of distance from each other, and from other substituted Xyl. (Methyl)glucuronidated and pentosylated Xyl were also found acetylated, and therefore, all substitution patterns determined in callus xylan were 'in frame' with each other.

Supporting this, the MS analysis of GH11 xylanase products showed that [Me]GlcA and Pent in callus xylan are spaced by two backbone Xyl, and no contiguous substitutions or three consecutive undecorated Xyl were detected. Neighbouring [Me]GlcA and Pent substitutions had a predominant arrangement, with the Pent at the RE side of [Me]GlcA substitutions (~U$^{[me]}$XN~ or ~PXN~). However, minor products also exhibited an inverted order (~NXU$^{[me]}$~; not shown). In addition, GH11 did not produce XNXX or XNXNXX structures, suggesting there are no Pent substitutions isolated from GlcA modification on callus xylan. The production of XU^{me}XU^{me}XX by GH11 indicates the neighbouring of two MeGlcA substitutions. In summary, GH11 produced oligosaccharides that only had a spacing of two Xyl between modifications.

The even pattern of substitutions on primary-wall xylan suggest they may have a role in interacting with cellulose. Ac are found in secondary cell walls predominantly every other backbone Xyl (Chapter 4; Busse-Wicher et al., 2014; Chong et al., 2014). The molecular dynamics data reported by Busse-Wicher et al. (2014), indicate that alternated backbone substitutions allow xylan to adopt a two-fold screw configuration and interact with cellulose microfibrils in the secondary wall. The interaction of two-fold xylan with cellulose was later experimentally confirmed by solid-state NMR by Simmons et al (2016). Their data shows that *irx3*, which is deficient of cellulose in

secondary cell walls, has a small signal of two-fold xylan in the proximity of cellulose. The authors explained that this vestigial signal is possibly primary cell wall contamination in the sample. Their data, however, implies that there is xylan-cellulose interaction in primary cell walls, which supports the view that primary-wall xylan interacts with cellulose and the importance of the compatible pattern of primary-wall xylan substitutions, shown in this chapter. Further investigation via solid-state NMR on ^{13}C-enriched callus would confirm whether this is indeed the case.

5.3.2 Arabinosyl and *O*-acetyl substitutions on primary-wall xylan are found in the same backbone residues, suggesting the even pattern is important

The third aim of this chapter was to describe the acetylation pattern of primary-cell-wall xylan. As two different types of xylan were found, both substitution patterns were investigated. The MS analysis of GH30/GH3 products on callus and stem primary-wall xylans determined that both polymers are acetylated. As discussed above, in both the acetylation follows an even pattern which would allow them to interact with the hydrophilic surface of cellulose in a two-fold screw. However, the patterns of acetylation exhibited by both xylans was different. A diagram with the two proposed acetylation patterns is presented in Figure 5.12, and to explain the difference between the two patterns, the possible Ac positions on the backbone is next explained in detail.

Stem primary-wall xylan showed three acetylated Xyl on the backbone in the repeating xylohexaose. The first acetylated position is the Xyl that carries the MeGlcA or PUX substitution (Xyl-2), which was found to be acetylated in all structures determined. The position of this Ac is consistent with the reported by Chong et al. (2015) on the Xyl carrying a PUX structure. The second acetylated residue was Xyl-4, which was predominantly found di-acetylated. The third acetylated residue was Xyl-6, which was found either mono-acetylated, or non-acetylated.

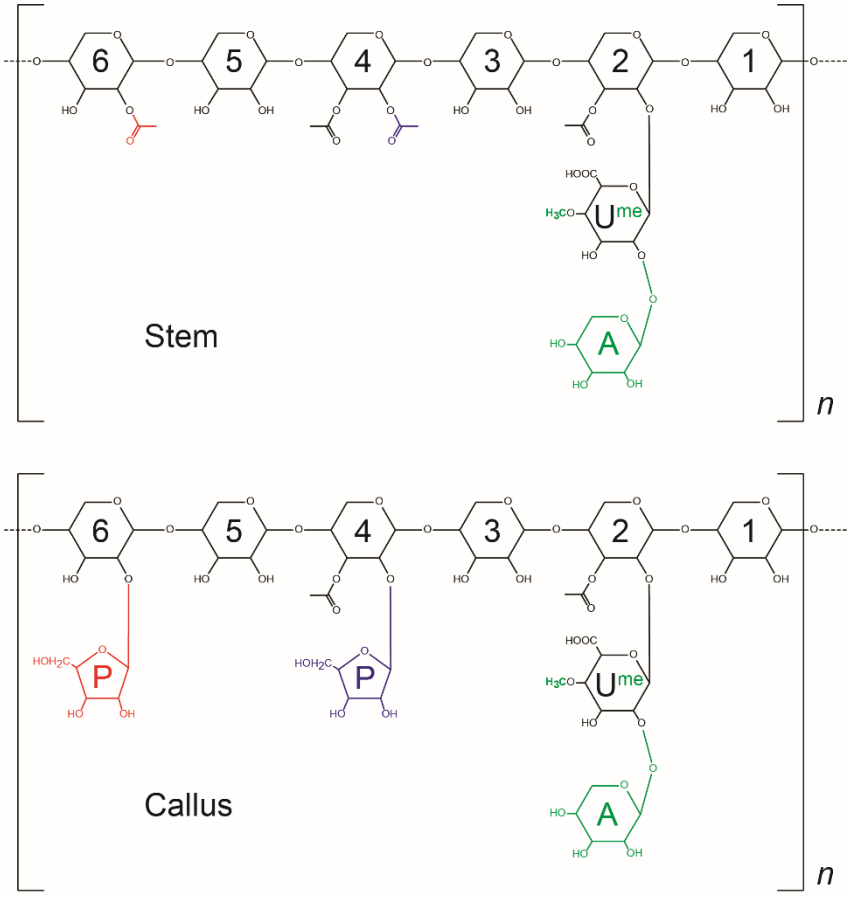


Figure 5.12: Models of primary-cell-wall xylan substitution patterns in *Arabidopsis*. These models consist on repeating units of xylohexaose with every-other Xyl substituted. Structures in black are constant. Xyl in black are numbered 1 – 6 from the RE. The acetyl substitution in blue is present predominantly. Side chains in red are also predominant but could be absent in minor occurrences. Pentosyl substitution in blue can be present, absent or replaced by an acetyl group. The linkages of the pentosyl and Ac substitutions to the backbone Xyl were not determined, but for simplicity they were drawn as either O-2 or O-3. Pent substitutions (P) are shown as Ara*f* but it was not determined whether they were arabinose or xylose in some cases. Structures in green are normally mutually exclusive. Only in minor cases both the methyl group and Ara*p* (A) can be present in a single GlcA modification on callus.

Callus xylan exhibited only two acetylated Xyl, however, also some Pent decorations were found in this type of xylan. The Xyl carrying the MeGlcA or PUX modification was always found acetylated (Xyl-2). The second acetylated residue was Xyl-4, which was found either mono-acetylated, di-acetylated, or co-substituted with an Ac and a Pent. In contrast to stem xylan, callus xylan Xyl-6 was not found acetylated, however, a pentosyl substitution was found on this backbone residue in most structures determined.

The similarity in the positions between Pent in callus xylan and Ac in stem primary-cell-wall xylan suggests that the two substitutions may have a similar role. Perhaps the synthesis of pentosyl substitutions on the callus xylan may be replaced by Ac on the stem xylan, which could be caused by a different sugar contents in the growth conditions. This hypothesis agrees with Gille and Pauly (2012), who explained that adding *O*-acetyl-substituents to wall polysaccharides instead of sugar residues is energetically favourable for the plant, as less carbon is deposited in the wall. They elaborated this hypothesis to explain the difference in xyloglucan substitutions between eudicots and grasses. However, the difference found for xylan substitutions in this work, varied within a single species.

Arabinosyl xylan sidechains are not very frequent in eudicots (Scheller and Ulvskov, 2010). Some callus xylan GH30 products showed sensitivity to GH62 arabinofuranosidase, suggesting the presence of Ara*f* on the backbone. In agreement, Tan et al. (2013) showed a 2-linked arabinosyl sidechain on xylan from *Arabidopsis* cell suspension cultures. They mention they found a 3-linked Ac on the arabinosylated Xyl, but it is not explained by the data shown in the report. However, this is also consistent with detected backbone Xyl co-substituted with Ara*f* and Ac. 2- and 3-linked Xyl sidechains have also been reported in psyllium husk xylan (Kennedy et al., 1979; Samuelsen et al., 1999), which contradicts the evidence of Pent being Ara*f*. However, reports of Xyl sidechains in eudicot xylan are mostly narrowed to the genus *Plantago*. In *Arabidopsis* only one GT61 xylan xylosyltransferase (Voiniciuc et al., 2015) has been proposed. Their evidence which suggests MUCI21 is a xylosyltransferase is based on a monosaccharide and a linkage analyses of *muci21* mucilage. However, no activity has been tested, therefore, their evidence for a eudicot GT61 protein acting as xylosyltransferase is not very strong.

To confirm the identity of the putative Ara*f* found in callus, it would be useful to analyse the xylan by solution-state NMR, and investigate whether these side chains are sensitive to the inactivation of the reported GT61.

5.3.3 Birchwood phloem and xylem xylan is structurally equivalent to different *Arabidopsis* stem primary-wall xylan

The xylan in the top (newly grown) and bottom (mature) portions of *Arabidopsis* stems differ in their [Me]GlcA substitution patterns. The reason of this difference has been attributed to the enrichment for primary or secondary cell walls respectively (Mortimer et al., 2015). However, to understand whether this enrichment is caused by a timing effect on the growth or due to tissue specificity of the walls, one aim of this chapter was to determine whether the xylan of xylem and phloem have different structure. Fascinatingly, the similarity of the [Me]GlcA substitution pattern found in birch phloem to the *Arabidopsis gux1/2* stems support the view that phloem cells have thickened primary walls. The function of the xylan in phloem and xylem may be therefore correlated with the function of the tissue. To explain the similarities found between birch and *Arabidopsis* xylan, the [Me]GlcA substitution pattern of xylan from *gux1/2* young stems is next discussed firstly, secondly it is compared to GUX1- and GUX2-mediated [Me]GlcA patterns on secondary-cell-wall xylan, and thirdly, birch phloem and xylem xylans are discussed.

The [Me]GlcA substitution pattern of *Arabidopsis* stem primary-wall xylan is even and with fixed spacing. GH30 glucuronoxylanase products of *gux1/2* stem primary-cell-wall xylan in this thesis showed on PACE two predominant species, each with six backbone residues; one of them was $U^{[me]}X_6$ and the other PUX_6, and in agreement with this, both species were produced in a similar GH30 hydrolysis analysis published by Mortimer et al. (2015). The MS analysis of GH30/GH3 products on acetylated xylan, on the other hand, revealed that most GlcA substitutions on GH30 products were further modified, either with a methyl or a pentosyl groups, and the amount of unsubstituted GlcA was only just above the detection limit. The two further modifications found on the GlcA in GH30 products also agree with the published MS analysis of GH11 products on *gux1/2* young stems (Mortimer et al., 2015). The [Me]GlcA substitution pattern in *Arabidopsis* primary cell walls is, therefore, essentially different to the pattern of stem secondary-cell-wall xylan.

Both substitution patterns on *Arabidopsis* stem secondary-cell-wall xylan have multiple spaces between [Me]GlcA backbone modifications, and both have some similarities to the one found on primary cell wall (Chapter 4; Bromley et al., 2013). It was

previously discussed that GUX2 produces predominant spaces of 5, 6 and 7 backbone Xyl between [Me]GlcA. These spacings are more frequent than the ones that GUX1 produces, which are of 8, 10 and 12, predominantly. On the other hand, the [Me]GlcA transferred by GUX1 are spaced one to another by an even number of backbone residues. GUX3 decorations on primary-wall xylan combine two characteristics from the patterns in secondary-wall xylan; the [Me]GlcA substitutions are at short distances (as in GUX2 pattern), and they are spaced by even number of backbone residues (as in GUX1 pattern). However, the acetylation context, and the further modification of the [Me]GlcA with Ara*p* (P) makes primary-wall xylan unique.

Birch phloem and xylem xylan exhibited contrasting GlcA substitution patterns. In birch phloem xylan, the MeGlcA substitutions were constantly spaced by six backbone residues. The two predominant GH30 products on phloem xylan co-migrate with $U^{[me]}X_6$ and PUX_6 from *gux1/2* stem primary-wall xylan. A smaller proportion of the MeGlcA substitutions on phloem xylan carry an Ara*p* substitution. Evidence for this is shown by the peak intensity of $PU^{me}X_6$ on MS spectrum which is lower than of $U^{me}X_6$, and supported by a similar trend in the densitometric intensity on PACE of both oligosaccharides (not shown). This proportion found in birch phloem is consistent with the lower proportion of arabinosylated to non-arabinosylated [Me]GlcA in *gux1/2* stem primary-wall xylan. On the other hand, birchwood xylem xylan exhibited a GlcA substitution pattern related to secondary cell walls. GH30 hydrolysis of xylem xylan showed on PACE the production of species with multiple numbers of backbone residues, and a predominance for products with even number of backbone Xyl. This is consistent with the pattern observed in *Arabidopsis* secondary cell walls (Chapter 4) and the reported in aspen (Derba-Maceluch et al., 2015).

The stem tissues studied in this chapter have great similarities between species. In comparison, birch phloem xylan GlcA substitution pattern has great similarity to the one found in *gux1/2* young stems. Possibly, the enrichment of primary cell walls observed in newly grown stems in *Arabidopsis* may be due to a temporary effect, in which primary- and secondary-cell-wall-enriched tissues (like phloem and xylem) exist in similar quantities. Then in later stages, secondary cell walls from hardwoods thicken by producing high amounts of cellulose, lignin, and secondary cell wall patterned xylan. Given the similarity between primary-cell-wall and phloem xylan, it would be expected to find a similar Ac substitution pattern of phloem xylan.

The results here obtained provide with information towards a better understanding of the different types of cell walls, and components in woody plants. It would be interesting to determine the xylan acetylation pattern in the different tissues.

5.3.4 GH30 glucuronoxylanase from different species have distinctive specificity to primary-cell-wall xylan substitutions

The GH30 glucuronoxylanase requirement of [Me]GlcA modifications on the substrate was an effective tool for describing the xylan substitutions pattern. The backbone size of GH30 products rely on the distance between the [Me]GlcA substitutions on the xylan substrate, unlike other xylanases such as GH10 or GH11 which require the xylan to have an unsubstituted gap. The hydrolysis of non-acetylated xylan by GH10 or GH11 produce smaller oligosaccharides than GH30 (Chapter 6; Chong et al., 2014), but when the xylan is acetylated mostly on every other backbone residue, GH11 and GH10 accessibility to the backbone is limited. In contrast, the efficiency of GH30 was not severely affected. In this way, having more [Me]GlcA substitutions on acetylated xylan allows GH30 to access it more than the other xylanases. Furthermore, in agreement with Mortimer et al. (2015), stem primary-wall xylan [Me]GlcA decorations are mostly spaced by six backbone residues. Therefore, GH30 products are long enough to describe a substitution pattern, and short enough to detect them by MS. Also, the GH30 requirement of a GlcA on the subsite -2, restricts the position of the [Me]GlcA to the second Xyl from the RE.

GH30 glucuronoxylanase enzymes from different organisms showed different sensitivity to the modification of GlcA with Ara*p*. Primary-wall xylan from *Arabidopsis* stem and root are occasionally further modified with Ara*p* (Mortimer et al., 2015; Peña et al., 2016). The described young stem xylan from *gux1/2* was used as substrate to test the specificity of three GH30 from different bacteria. The products of BoGH30 (*Bacteroides ovatus*), EcGH30 (*Dickeya dadantii*) and CtGH30 (*Clostridium thermocellum*) observed by PACE differed in their [Me]GlcA modifications. BoGH30 main products were $U^{[me]}X_6$ and PUX_6. This is a consistent reproduction of the BoGH30 activity reported by Mortimer et al. (2015) with a similar substrate. Interestingly, PUX_6 was not a predominant product by EcGH30 or CtGH30 when hydrolysing *gux1/2* xylan. In addition, EcGH30 and CtGH30 produced some longer oligosaccharides of

approximately double the size of $U^{[me]}X_6$ or PUX_6, which could be generated from a polysaccharide with a PUX sidechain flanked by two cleavable sites with $U^{[me]}$ substitutions.

Structural data for EcGH30 shows that Tyr-255 forms a hydrogen bond with the 2-hydroxyl group on the MeGlcA (Urbániková et al., 2011). *In silico* analysis shows that this tyrosine is well conserved in the GH30-8 family and is present in CtGH30 (Louis Wilson, personal communication). However, the corresponding residue on BoGH30 is leucine. This could explain the difference in specificity, in which tyrosine poses an obstacle to Ara*p* binding that leucine does not. A phylogenetic analysis of CAZy GH30 sequences shows that BoGH30 belongs to different subgroup to EcGH30 and CtGH30 (Louis Wilson, personal communication).

These preliminary results on the substrate specificity of GH30 are worth investigating further for developing enzymes that could digest plant-based material more effectively.

Chapter 6: Finding putative xylan acetyltransferases

6.1 Introduction

The distribution of Ac in several distinct domains of xylan was described in the previous two chapters. Xylan from primary and secondary cell walls showed different environments around Ac side chains, such as GlcA with further modifications and Ara*f* substitutions. The pattern of Ac in the compatible and incompatible domains of *Arabidopsis* xylan in secondary cell walls was shown to be different. The compatible domain is decorated with alternating acetylated Xyl and unsubstituted Xyl, and in the incompatible domain, the pattern was shown to have a dominance for consecutive acetylated Xyl (Chapter 4). Xylan from primary cell walls showed consistent patterns of acetylation, in which acetylated Xyl are alternated with unsubstituted Xyl (see Figure 1.21). The backbone Xyl can be mono-substituted with Ac at O-2 or O-3, di-acetylated at O-2 and O-3, or co-substituted with [Me]GlcA at O-2 and Ac at O-3 (Scheller and Ulvskov, 2010). In Chapters 4 and 5, *Arabidopsis* xylan backbone Xyl were also shown to be co-substituted with Ac and a pentosyl sidechain. The different positions and configurations of Ac on xylan may have different functions. Some of these functions may be associated with maintaining the two-fold and three-fold-screw structure of xylan, or in facilitating the further methylation of GlcA substitutions.

The synthesis of such complex distribution of Ac on different types of xylan suggests that the process of xylan acetylation requires a highly regulated machinery. Different acetyltransferase activities are possible, because the acceptor and the Ac regiospecificity are different in each case. Therefore, to investigate the process of xylan synthesis, it is important to identify different xylan acetyltransferases.

The TBLs are a family of predicted acetyltransferases active on different carbohydrates, including xylan (Bischoff et al., 2010a). This family is composed of 46 members in *Arabidopsis*, and they all share sequence identity. The TBLs have an N-terminal transmembrane domain and two globular domains. The domain at the C-terminus (DUF231) has a conserved DxxH motif. The domain in the middle was named the TBL domain and contains a conserved GDS motif. The DxxH and GDS motifs are present and conserved along esterases and lipases (Akoh et al., 2004; Bischoff et al., 2010a; Gille and Pauly, 2012), which is why they are predicted to have

acetyltransferase activity. The following are a few examples of TBLs that have been identified to be involved in the biosynthesis of cell wall polysaccharides: TBR and TBL3 are associated with the correct cellulose synthesis and deposition into the cell wall, yet the cellulose content does not vary in *tbr* and *tbl3* mutants (Bischoff et al., 2010a); TBL22/AXY4-L and TBL27/AXY4 were shown to be involved in xyloglucan acetylation (Gille et al., 2011; Zhu et al., 2014); TBL23, TBL24, TBL25 and TBL26 can transfer Ac to mannohexaose (Zhong et al., 2018c); TBL10 was recently shown to be involved in the acetylation of pectin (Stranne et al., 2018); and ESK1 (TBL29/XOAT1), which at the beginning of this work, was the only TBL known to be responsible for part of xylan acetylation.

ESK1 is a Golgi-localised xylan-specific acetyltransferase (Urbanowicz et al., 2014; Yuan et al., 2013). The Acs transferred by ESK1 have been shown to be initially on O-2 on the backbone Xyl, and then they spontaneously and reversibly migrate between positions O-2 and O-3. This implies that ESK1 is stereospecific for the position O-2. The *esk1* mutant is dwarfed, has collapsed xylem vessels, and has around 40 – 50% of xylan *O*-acetyl groups compared to the WT (Xiong et al., 2013). Because *esk1* xylan is still partly acetylated, it implies that ESK1 is not the only xylan acetyltransferase. At the start of the work in this chapter, the roles of other TBLs was unclear. The remaining Ac on *esk1* xylan are mostly nearby [Me]GlcA substitutions, and traces of isolated Ac on the backbone have been reported, however, in *esk1* the main acetylation pattern of xylan is lost (Figure 1.21; Grantham, 2016). The inactivation of ESK1 also affects the downstream process of xylan glucuronidation, altering the [Me]GlcA substitution pattern of xylan by GUX1 (Grantham et al., 2017).

Apart from the TBLs, other proteins in the xylan acetylation pathway have also been identified. RWA proteins have been suggested to act as Ac-CoA (acetate donor) transporters into the Golgi apparatus, and the *rwa* mutants show deficits in xylan and xyloglucan acetylation (Gille and Pauly, 2012; Grantham, 2016; Manabe et al., 2013). However, in contrast with *esk1*, in *rwa* the GlcA pattern of xylan is essentially unaffected (Grantham et al., 2017). Downstream of RWA, AXY9 has been hypothesised to use the internalised Ac-CoA to produce an acetylated intermediary that is substrate for the TBLs (Schultink et al., 2015). The *axy9* mutant has a severe growth phenotype and its acetate content is reduced by 67% in pectin from seedlings and 80% in total xylan with respect to the WT (Schultink et al., 2015). AXY9 may

produce the substrate for all TBLs, or it could be selective for certain acetyltransferase activities. Alternatively, AXY9 could acetylate the xylan directly. Therefore, its function is worth investigating further.

Downstream of the xylan acetylation process other enzymes decorate the xylan with additional substitutions. GUX1, GUX2 and GUX3 substitute the backbone with GlcA, and then GXMs methylate the GlcA substitutions. Different acetylation mutants have different effects on the methylation of the GlcA. In WT xylan around 60% of the GlcA are methylated, in *esk1* the proportion is increased (~67%), and in *rwa* the proportion is decreased (~50%) (Grantham, 2016). This implies that the position of the Ac transferred by ESK1 and by the other acetyltransferases impact differently on the downstream xylan decoration processes. Thus, these Ac have different functions.

The process of xylan acetylation is therefore not completely understood. Several xylan acetyltransferase activities with different regiospecificity are unresolved. The aim of this chapter was to identify acetyltransferases additional to ESK1 that would be also acting on xylan. In order to do this, the main aim was subdivided into five sub-aims, which is next explained.

First, putative acetyltransferases acting on xylan were identified from the TBL family, by implementing phylogenetic analysis of the TBL family. Because ESK1 activity is specific for xylan, ESK1-paralogues with high amino-acid similarity have higher probability of xylan specificity. However, it was also possible that the high similarity to ESK1 would define the O-2 regiospecificity of the transferase activity, and O-3 acetyltransferases may group in another branch of the phylogenetic tree. To evaluate this, the conservation of the genes in the ESK1 clade in other higher plants was analysed by constructing a phylogenetic tree with other species. To determine which putative acetyltransferases may be active on primary- or secondary-cell-wall xylan synthesis, the expression patterns of the closest paralogues to ESK1 were analysed, and compared to the patterns of GUX1, GUX2 and GUX3.

Second, to determine if any other TBLs have the same activity as ESK1, they were expressed under the ESK1 promoter in the *esk1* mutant. Then, if they complemented the mutant, this would indicate they have similar biochemical activity but a different expression pattern insufficient to be redundant in *Arabidopsis*.

Third, to investigate the possible activity of the TBL in the ESK1 clade, knockout insertional mutants were isolated, and the lack of expression of the TBL genes in the mutants was verified. Because possible gene duplication events are likely to occur in *Arabidopsis* some TBL genes may be redundant. To overcome the redundancy problem, the *tbl* lines were inter-crossed to generate mutants lacking both putative redundant genes.

Fourth, to screen for effects of the inactivation of TBL genes on the xylan, the substitutions and substitution patterns in the *tbl* mutants were evaluated by using several biochemical techniques. The inactivation of ESK1 produces an indirect alteration of the [Me]GlcA pattern of xylan, therefore, the impact of the inactivation of the TBL genes on the [Me]GlcA pattern of xylan was evaluated by PACE. Then, deficiencies in mono-acetylated and di-substituted Xyl were also investigated in the mutants by PACE using combinations of GH11 xylanase, CE4 acetylesterase and alkali. The identification of the deficient-Ac position in the *tbl* mutants' xylan was carried out by analysing the hydrolysis products by MS.

Fifth, the *tbl* mutants in which acetylation deficits were found were investigated further. To identify the position of the Ac deficiency, the *tbl* mutants' xylan was digested with a xylanase and the products were analysed by MS. The effect of the acetylation deficit on the methylation of the xylan GlcA substitutions were analysed by DASH as previously described (Li et al., 2013). Additionally, to confirm, the position of the deficient Ac, and the effect on the GlcA methylation, the mutants' xylan was analysed by NMR spectroscopy.

The work presented in this chapter elucidated two TBLs (TBL32 and TBL33) required for the 3-*O*-acetylation of glucuronidated Xyl, one TBL (TBL28) that has a similar activity to ESK1 (TBL29), and one TBL (TBL3) which is involved in secondary-cell-wall xylan acetylation.

6.2 Results

6.2.1 In silico search for putative acetyltransferases acting on xylan

6.2.1.1 ESK1 groups with a conserved TBL phylogenetic clade

The TBL family is composed by 46 members which are predicted to be carbohydrate acetyltransferases. To evaluate which may be acting on xylan, a phylogenetic analysis was carried out. A phylogenetic tree of the TBL family was previously reported by Bischoff et al (2010a) which shows that ESK1 groups with eight additional proteins. The *Arabidopsis* TBL tree shown in Figure 6.1 was re-generated using a different software from the publication, but the topology and branch lengths are consistent with the previously reported. To construct the tree, the amino acid sequences of the 46 *Arabidopsis* members of the TBL family were obtained from TAIR and aligned using the MUSCLE algorithm. A maximum likelihood (ML) phylogenetic tree was constructed using MEGA7 software following the default instructions published by Hall (2013), and applying a bootstrap (*n* = 1000 trials).

ESK1 grouped with eight other TBL protein sequences (red branch). This branch encompasses TBL3 and TBL28 to TBL35, with a bootstrap value of 100, which is indicative of the robustness of the grouping. The purple branch shows the closest homologue to the red branch (TBL36) which was also included in some of the later analyses due to the high level of expression of TBL36 (see section 6.2.1.3). The divergence towards the end of the branches in the tree suggests that some *TBL* genes may have been duplicated relatively recently, which is consistent with the levels of gene duplication in the *Arabidopsis* genome (Panchy et al., 2016). In this way, it is likely that TBL32 and TBL33 may have a similar biochemical activity. TBL34 and TBL35 could also be redundant, as well as TBL28 and ESK1. The locus of *TBL30* (*AT2G40160*) is in tandem with the locus of *TBL28* (*AT2G40150*) in *Arabidopsis* genome, which suggests they share a common ancestor. Also, because TBL28 appears to be closer to ESK1 in the tree, this suggests that *TBL30* may be a pseudogene. It is possible that all ESK1-homologues in the red clade are redundant to ESK1 and/or may have a similar enzymatic activity. Alternatively, the ESK1 closest

homologues may differ in their activity, but they group together because they all may act on xylan.

The TBLs in the rest of the tree (in black) may be associated with the acetylation of other carbohydrates. Some of the branches in the tree have been labelled also with their alternative names, corresponding to what has been published about them. In this way, for example, close homologues to TBL22/AXY4-L and TBL27/AXY4 could be associated with xyloglucan acetylation, and pectin could be acetylated by the TBL homologues near TBL10.

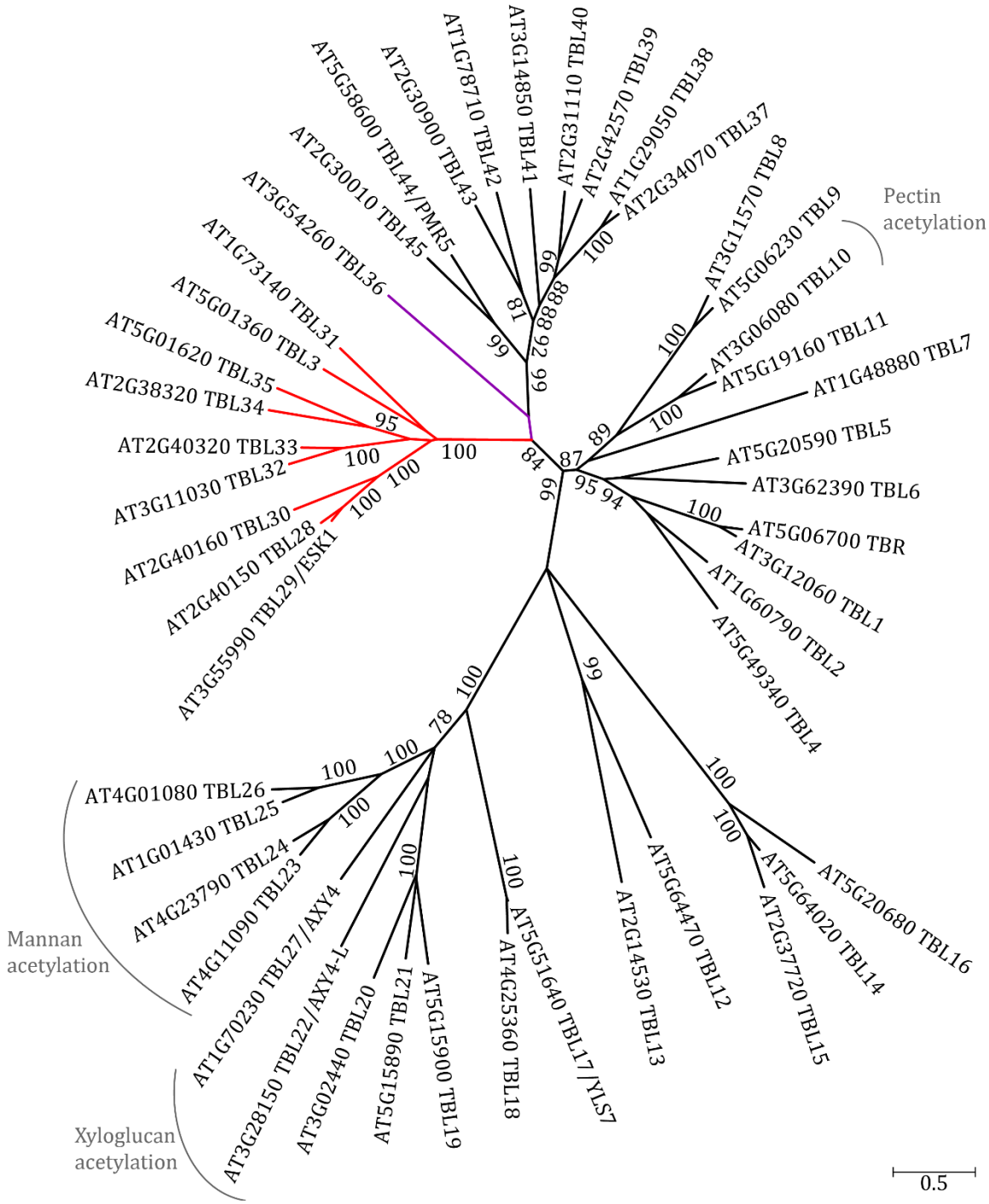


Figure 6.1: *Arabidopsis* TBL protein family phylogenetic tree.
The 46 *Arabidopsis* TBL amino acid sequences were obtained from TAIR and aligned using the MUSCLE algorithm. A maximum-likelihood phylogenetic tree was constructed using MEGA7 software following the default instructions published by Hall (2013), and applying a bootstrap (*n* = 1000 trials). In red is highlighted the clade with the xylan-specific acetyltransferase ESK1/TBL29 and its closest homologues. The purple branch highlights the closest homologue to the red branch, with relatively high levels of expression (see Figure 6.4B). Only bootstrap values over 75 are shown. Possible functions of reported genes related to the acetylation of other polysaccharides is shown in grey and described in the text.

6.2.1.2 Conservation of ESK1 closest homologues in higher plants

Conservation of the sub-clades might indicate a diverged function needed in all the plants. To determine if the genes from the TBL family are recent *Arabidopsis* duplications or they are conserved, a phylogenetic tree was built including other representative species in higher plants. A scheme of the diversification of higher plants is shown in Figure 6.2, which shows the relationship between higher plants, angiosperms and gymnosperms. Because the Gene-Ontology-enriched genomic data of the species shown in red were available in the PLAZA database (Van Bel et al., 2012), they were included in the phylogenetic analysis of the TBL family.

The xylan acetylation is different in the different plant clades (Busse-Wicher et al., 2016b), and therefore it is expected to find differences in the acetyltransferase activities. Conserved acetyltransferase functions are expected to group together, and separated from other sub-cades. Monocot xylan has less acetate groups and some sugar sidechains that are different to eudicots (Feijão, 2016), therefore, possibly some ESK1 closest homologues from *Arabidopsis* may not group with monocot members. In contrast, gymnosperm xylan has undetectable levels of acetate (except for gnetophytes) (Busse-Wicher et al., 2016b), which makes the gymnosperm TBLs a negative control for xylan acetyltransferases.

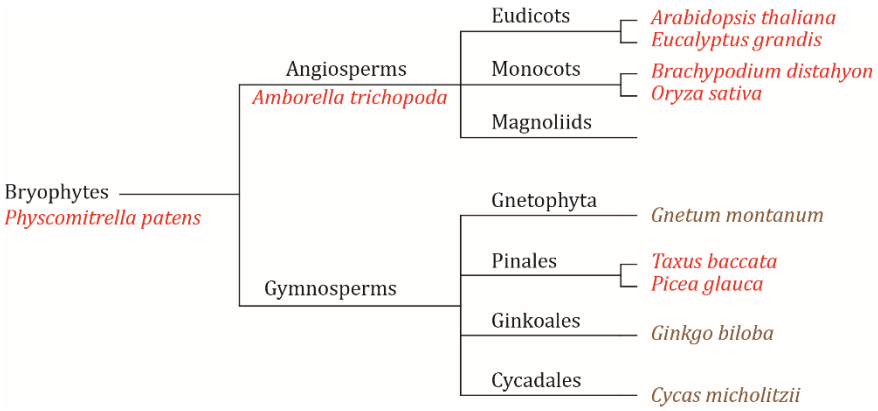


Figure 6.2: Schematic representation of the phylogenetic relationship between gymnosperm and angiosperm species.
Names in black show the phylogenetic clades. Transcriptomic data from the representative species (in red and brown) was used for the phylogenetic analysis of the TBL family. Species in brown are poorly annotated but were the only available in the PLAZA database for their clade. Species in red are have a relatively reliably annotated in the database, and the annotation of species shown in brown are relatively poor. Modified from Busse-Wicher et al (2016b).

To construct the tree, the ESK1 amino acid sequence from *Arabidopsis* was used to perform a BLAST search on PLAZA (https://bioinformatics.psb.ugent.be/plaza). A total of 429 sequences from the species in Figure 6.2 (including the 46 sequences of the *Arabidopsis* TBL family) were aligned using the MUSCLE algorithm. Highly variable, gapped N-terminal regions were removed, and a ML tree was constructed using FastTree software with a Shimodaira-Hasegawa robustness test (SH). The alignment and the tree are shown in the annex (Figure S4 and Figure S5). In this tree, 51 sequences grouped in the ESK1 clade (highlighted in red in Figure 6.1 and in Figure S5) with an SH p-value of 1, which indicates that this grouping is highly accurate. The sequence of AtTBL36 was added as outgroup (in purple), and the 52 sequences were used to build a smaller second tree. To construct a ML tree of only ESK1 clade, the 52 selected sequences were aligned using the MUSCLE algorithm (annex, Figure S6), the tree was built with MEGA7 software and it was bootstrap-tested with 1000 iterations (Figure 6.3).

To analyse the conservation of putative functions of the TBLs in *Arabidopsis*, the sub-clades of the higher-species ESK1-clade-tree is next described.

Arabidopsis TBL31 grouped with one sequence of *Eucalyptus*, and one from *Amborella*, with a bootstrap value of 100. Because *Amborella* is a common ancestor between eudicots and monocots, this grouping suggests a conservation of the function of TBL31 in angiosperms. Also, because no rice or *Brachypodium* sequences grouped in the TBL31 sub-clade, this suggests that grasses may have lost this function.

TBL3 grouped with one *Eucalyptus* sequence, one from *Amborella* and one from rice with a bootstrap value of 100. This suggests that the function of TBL3 may be conserved among angiosperms, including monocots.

ESK1 and TBL28 together grouped with one sequence of *Eucalyptus* and one from *Amborella* with a bootstrap value of 99. Possibly *ESK1* and *TBL28* originated from a recent gene duplication event in *Arabidopsis*, and their function may be conserved in eudicots.

The ESK1/TBL28 sub-clade grouped with a bootstrap value of 100 with 12 monocot sequences and TBL30. Interestingly, this suggests that the function of ESK1/TBL28 may have diversified in grasses. In a similar manner, proteins from the GT61 family

involved in xylan arabinosylation have also shown a broad diversification in grasses (Anders et al., 2012). It would be interesting to investigate any correlation between these two gene diversifications that may have occurred in grasses. The position of TBL30 was apart from ESK1 and not grouping with *Eucalyptus* sequences. Since TBL30 and TBL28 are contiguous genes in *Arabidopsis*, it is possible that TBL30 may have changed or lost its function which makes it more prone to generate changes in its sequence, which makes it group out of the ESK1/TBL28 sub-clade.

One sequence of *Amborella* grouped with TBL32 and TBL33, and three *Eucalyptus* sequences. This suggests that TBL32 and TBL33 may have the same enzymatic activity as each other but their conservation among eudicots suggests that this group may have different functions to other subclades. The grouping of three pairs of grass sequences in the TBL32/TBL33 sub-clade suggests that this enzymatic activity, although related, may have diversified in monocots.

TBL34 grouped with six *Eucalyptus* sequences with a bootstrap value of 98, which suggests gene duplication with differential expression, and that its function is conserved in eudicots. This sub-clade groups with one *Amborella* sequence and five grass sequences, suggesting that this function is conserved among angiosperms

TBL35 grouped with an *Amborella* sequence with a bootstrap value of 99, suggesting they may have a similar function. These two sequences grouped with the TBL34 subclade, with a bootstrap value of 93, suggesting they may have similar enzymatic activities, and diversified in angiosperms.

No gymnosperm sequences grouped in the ESK1 clade. Since fern xylan is possibly acetylated (A Li, unpublished data; Haghighat et al., 2016), this suggests that xylan acetylation may have diversified with the arrival of the angiosperms, and it may have been lost in most gymnosperms. The grouping of the nine ESK1 closest paralogs with six homologs in *Amborella* in Figure 6.3 suggests that between five and six different enzymatic activities are likely within the *Arabidopsis* TBLs. All nine ESK1 closest *Arabidopsis* homologues were selected to be studied in this work.

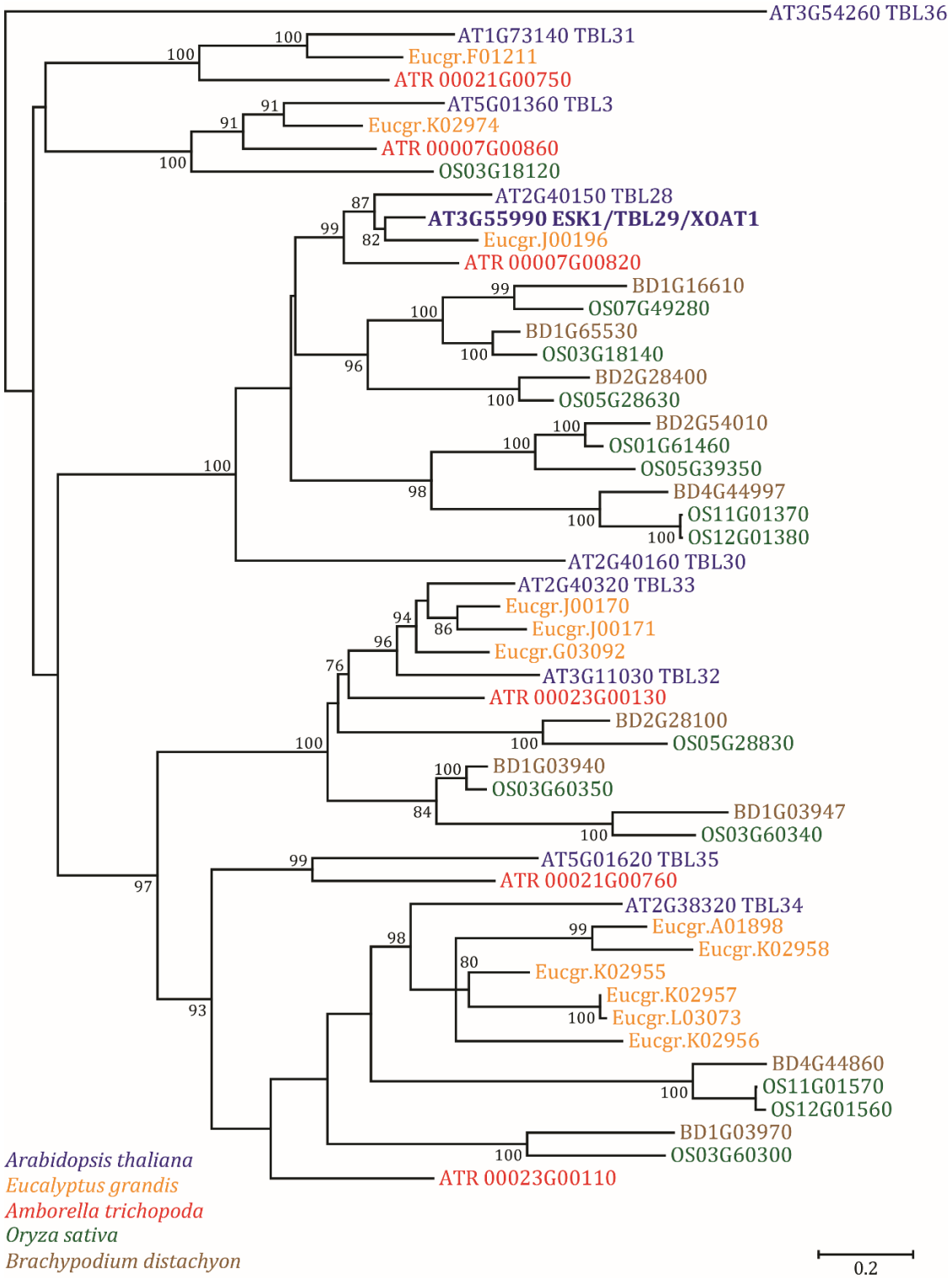


Figure 6.3: Phylogenetic analysis of ESK1 closest homologues in higher plants.
A first tree was constructed using ESK1 BLAST output sequences on the PLAZA database. The sequences were aligned using the MUSCLE algorithm (Figure S4), and the ML tree is shown in the annex (Figure S5). The sequences of the ESK1 clade (shown in red in Figure 6.1 and in Figure S5) were selected for the construction of the tree shown here, and AtTBL36 was added as outgroup. To build this tree, the 52 selected sequences were aligned using the MUSCLE algorithm, the tree was constructed using the software MEGA7 and bootstrap-tested with 1000 iterations. Only bootstrap values over 75 are shown.

6.2.1.3 Expression of *TBL* genes

The different xylans in *Arabidopsis* have been shown to be synthesised by different sets of glycosyltransferase enzymes, thus, the TBLs that add Ac may have a similar behaviour. Perhaps, the TBL functions differ in the xylan domains they acetylate, or alternatively, they may act on xylan from primary or secondary walls. To determine which TBL acts on which type of xylan, their expression levels and patterns were compared with GUX1 and GUX2 which decorate the compatible and incompatible domains of xylan respectively, and are expressed in secondary cell wall producing tissues (Chapter 4; Lee et al., 2012; Bromley et al., 2013). GUX3, whose function is specific to primary-cell-wall xylan (Chapter 5; Mortimer et al., 2015) was also included in the expression analysis.

TBL and GUX expression data were obtained from *Arabidopsis thaliana* Tiling Array Express (At-TAX, http://gbrowse.weigelworld.org/cgi-bin/attax), which possesses information of all ESK1 closest paralogues (Laubinger et al., 2008). Additionally, RNAseq expression data was extracted from TRAVA database which also has information of all ESK1 closest paralogues. Both data are presented in Figure 6.4; the graph with the mean normalised At-TAX values are shown in panel A, and the graph with RNAseq values is shown in panel B.

The data found in both databases exhibit a difference in reliability. Higher levels of expression of *GUX1* and *GUX2* were expected to be found in stem compared to other tissues, however this was not found in At-TAX (panel A). This shows At-TAX data is unreliable for this purpose, because it is inconsistent with what has been previously reported (Mortimer et al., 2010; Rennie et al., 2012), and therefore it was not used for further analysis. In contrast, in panel B, *GUX1* and *GUX2* show a typical differential expression in mature stem (internode). *GUX2* is also highly expressed in seeds. On the other hand, the expression of *GUX3* peaks in meristems (apex) and root. *GUX3* also seems to have a constitutive expression in the other tissues. The consistency of the *GUX1/2/3* RNAseq data to the previous publications indicates that the RNAseq values are more reliable than the array.

The RNAseq data shows that some of the analysed *Arabidopsis TBLs* are differentially expressed in stem, and others are expressed in a broader range of tissues. *ESK1*, *TBL3*, *TBL31*, *TBL33* and *TBL36* show a differentially high expression in stems

compared to their expression in other tissues. The expression pattern of the mentioned *TBLs* is similar to the exhibited by GUX1, suggesting that these TBLs may be involved in secondary-cell-wall xylan synthesis. In contrast, *TBL28*, *TBL30*, *TBL32*, *TBL34* and *TBL35* show a more spread expression pattern in the different tissues, without a clear predominance. Their expression pattern is therefore similar to the one of *GUX3*, suggesting that they may be related to primary-cell-wall xylan synthesis.

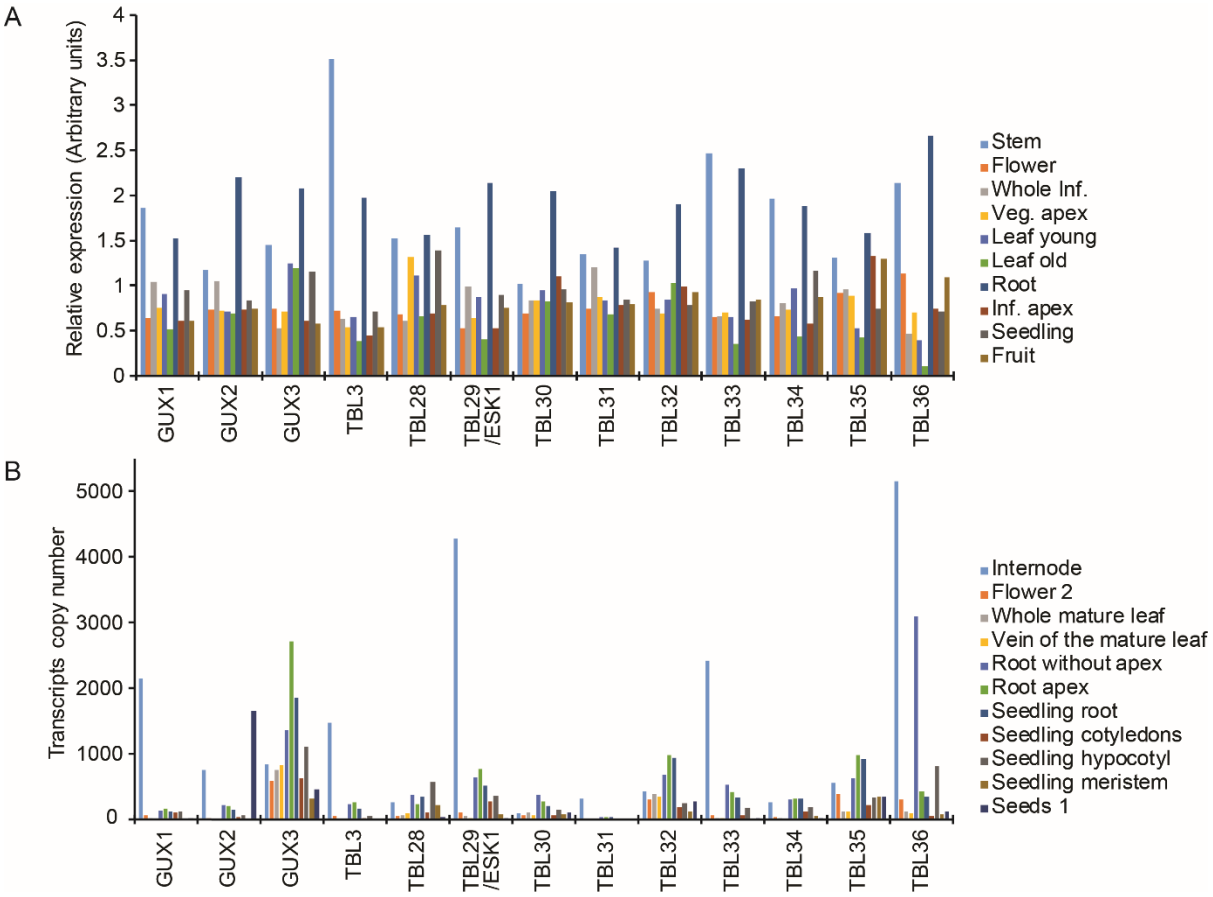


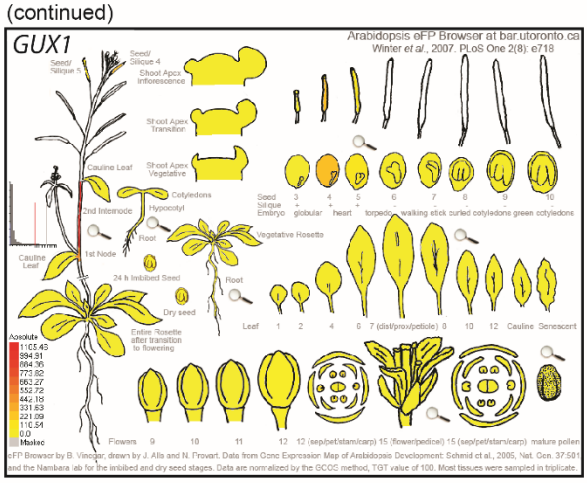
Figure 6.4: Expression levels of *ESK1* closest paralogues in different tissues.
A: At-TAX expression data from several tissues. The values were extracted and from Weigel World's TileViz (http://jsp.weigelworld.org/tileviz/tileviz.jsp) and mean-normalised per gene. B: RNAseq expression data acquired from TRAVA (http://travadb.org/).

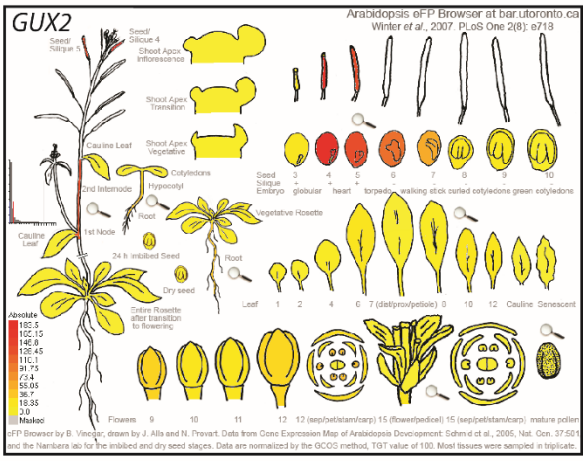
The two sources used for obtaining the TBL expression data produced two different outcomes, therefore it was decided to compare them with a third database. The usual expression database for *Arabidopsis* ATH1 does not have probes for *ESK1*, *TBL31* and *TBL33*, however, it does contain probes for the rest (Bassel et al., 2008; Winter et al., 2007). The ATH1 expression data from *GUXs* and some *TBLs* were obtained from eFP Browser and are presented in Figure 6.5. The expression pattern of *GUXs* and some *TBLs* in different developmental stages in a broad range of tissues is shown in panel A. Because some genes showed a differential expression in seeds, a detailed expression pattern in seeds is presented in panel B. To analyse the consistency of the data to the RNAseq, the expression patterns from ATH1 is next discussed.

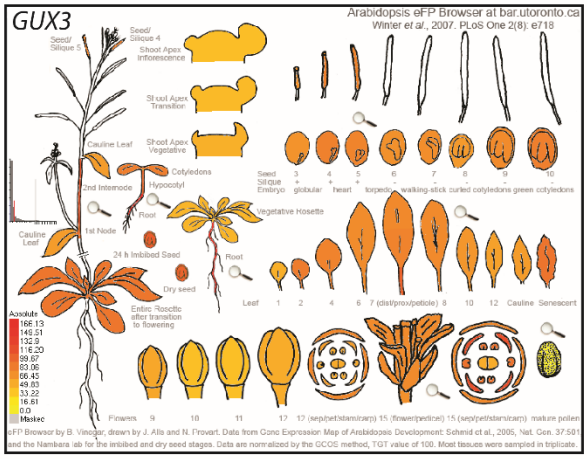
The maximum expression of *GUX1* is in stem, *GUX2* also has high levels in stem but its peak is in seeds, and *GUX3* is expressed widely. *TBL3*, *TBL34* and perhaps *TBL36* seem to have a similar expression pattern to GUX1, because their highest expression is in basal stem. In contrast, *TBL28*, *TBL32* and *TBL35* are expressed more ubiquitously following a similar pattern to *GUX3*. The expression predominance in stem for *GUX1*, *GUX2* and *TBL3* present in ATH1 is therefore consistent with the RNAseq data discussed above.

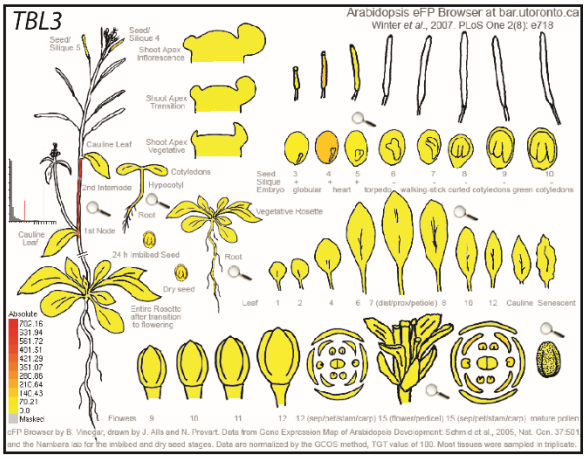
Two *TBL* genes have a peak of expression in seeds, similar to that exhibited by *GUX2* (Figure 6.5B). *TBL30* is almost only expressed in seed coat, and *TBL28* is expressed in seeds but also has basal expression levels in the rest of the plant. Perhaps *GUX2*, *TBL28* and *TBL30* may synthesise a particular type of xylan during seed development, which would be interesting to investigate further.

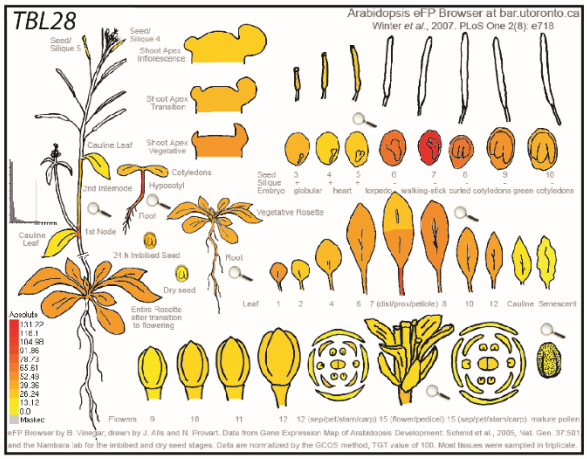
A (continued)

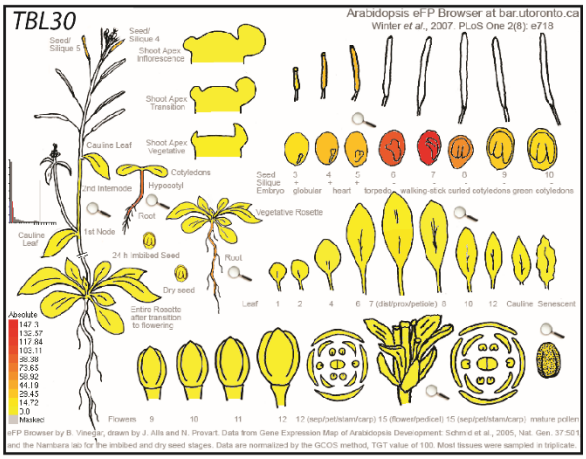

GUX1
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718

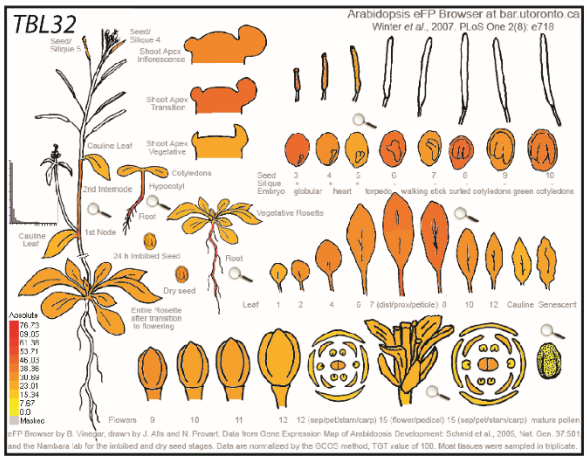

GUX2
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718

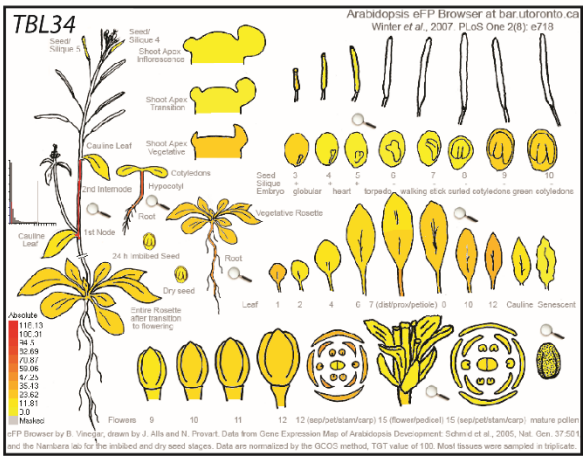

GUX3
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718


TBL3
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718

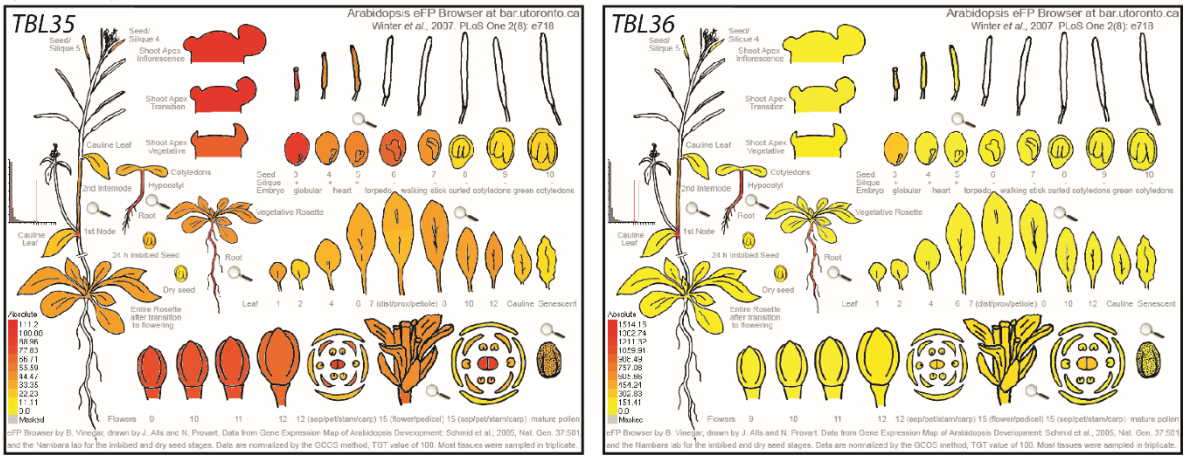

TBL28
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718


TBL30
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718


TBL32
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718


TBL34
Arabidopsis eFP Browser at bar.utoronto.ca
Winter et al., 2007. PLoS One 2(8): e718

A (continued)



B

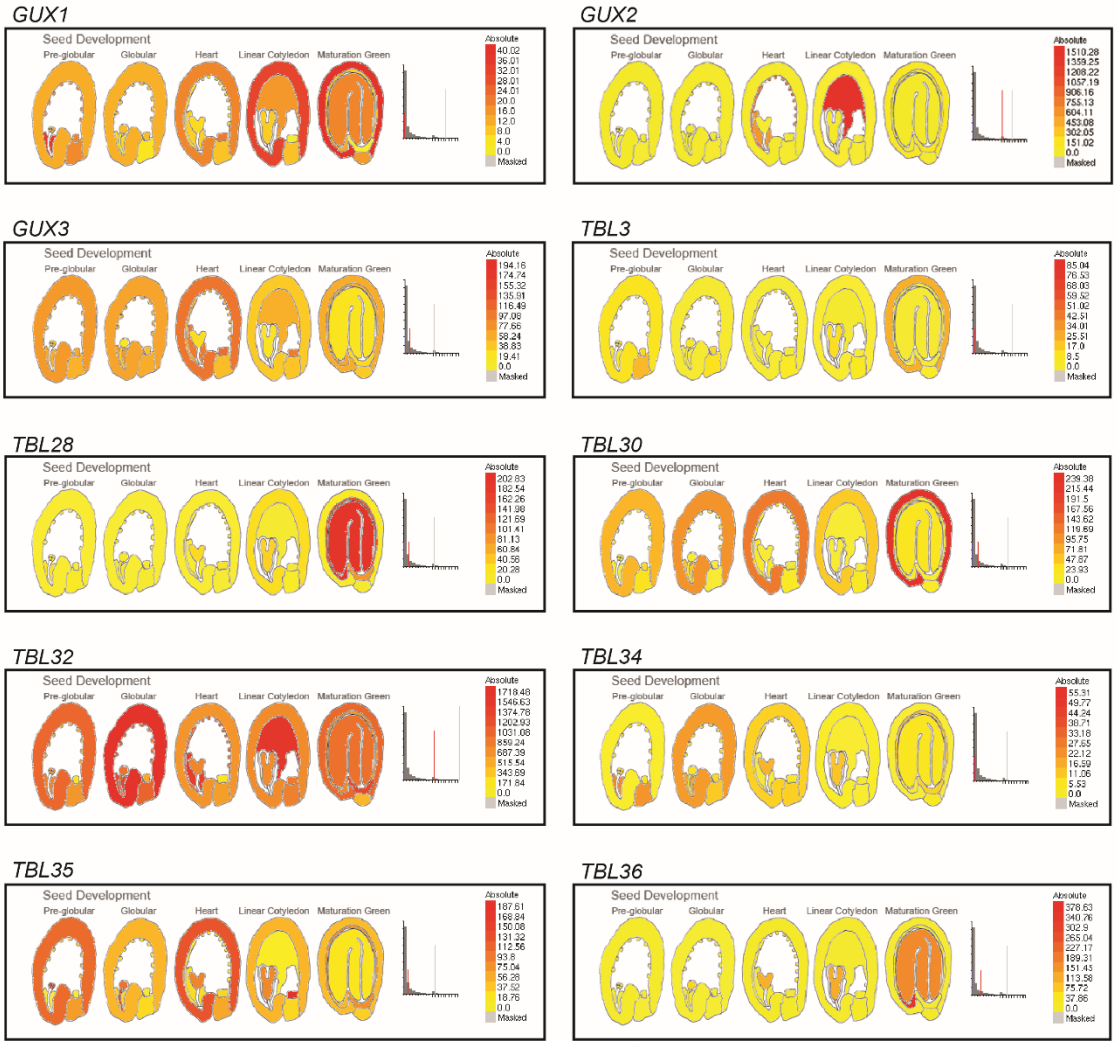


Figure 6.5: Heat-maps of *TBL* and *GUX* expression from ATH1 microarray.
The expression patterns of *Arabidopsis GUXs* and the *ESK1* closest paralogues in different developmental stages are shown. Expression levels on a broad range of tissues are shown in Panel A, while seed-specific expression is shown in panel B. The data were acquired from the eFP Browser website (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi).

6.2.2 Promoter swapped expression of *TBLs* can partially restore *esk1* phenotypes

It is possible that some TBLs may have a similar biochemical acetyltransferase activity to ESK1, yet not be redundant because of insufficient level or different tissue expression. If any TBL has similar activity, its expression in *esk1* could recover the growth to WT levels. It may also restore the compatible [Me]GlcA pattern of xylan, which is altered in *esk1* (Grantham et al., 2017). In order to investigate whether any of the ESK1 closest paralogues may have a similar enzymatic activity to ESK1, a promoter swap experiment was carried out. For this, *esk1* plants were transformed with vectors for the expression of the *TBLs* (listed in Figure 6.6) under the *ESK1* promoter, and the growth and the [Me]GlcA pattern of xylan in the transformed lines were analysed. The constructed vector, the height and [Me]GlcA substitution pattern of xylan in the produced lines are presented in the next three sections.

6.2.2.1 Construction of the expression vectors, and production of the promoter swapped *TBL* lines

In order to test whether ESK1 closest paralogues can restore the height and [Me]GlcA pattern of xylan, *esk1* plants were transformed with vectors for the expression of the *TBLs* under the promoter of *ESK1*. The promoter of *ESK1* would allow the *TBLs* to be expressed at similar levels and tissues where *ESK1* is normally expressed. To construct the vectors, the *TBL* full length coding regions and the *ESK1* promoter were individually synthesised and cloned in L0 vectors by Twist Bioscience (for Golden-Gate cloning see Chapter 2, section 2.3.2.3). The coding and promoter sequences were confirmed by DNA sequencing from the L0 vectors. The *TBL* genes were then fused to *ESK1* promoter, myc tag and Nos terminator on an L1 vector (transcriptional unit, not shown) by Golden-Gate cloning. Then, each of the genes was cloned into the final L2 vectors (Figure 6.6), which were transformed into *esk1-5* mutant plants via *Agrobacterium*. Because the expression of TBL3 in *esk1* produced an unexpected growth phenotype (shown in section 6.2.2.2), to investigate whether this effect is caused by a defect on xylan, Col-0 WT was also transformed with the vector for the expression of *pESK1:TBL3*. The transformed seeds were selected by seed fluorescence produced by the selection marker Oleosin-GFP. Up to 10 independent

lines were selected for each gene, then they were taken to T3 by selfing, and homozygous lines selected based on production of essentially 100% fluorescent seeds. No studies were carried out to confirm expression of the transgene in stem, however, the fluorescence in seeds indicated expression of the selection marker.

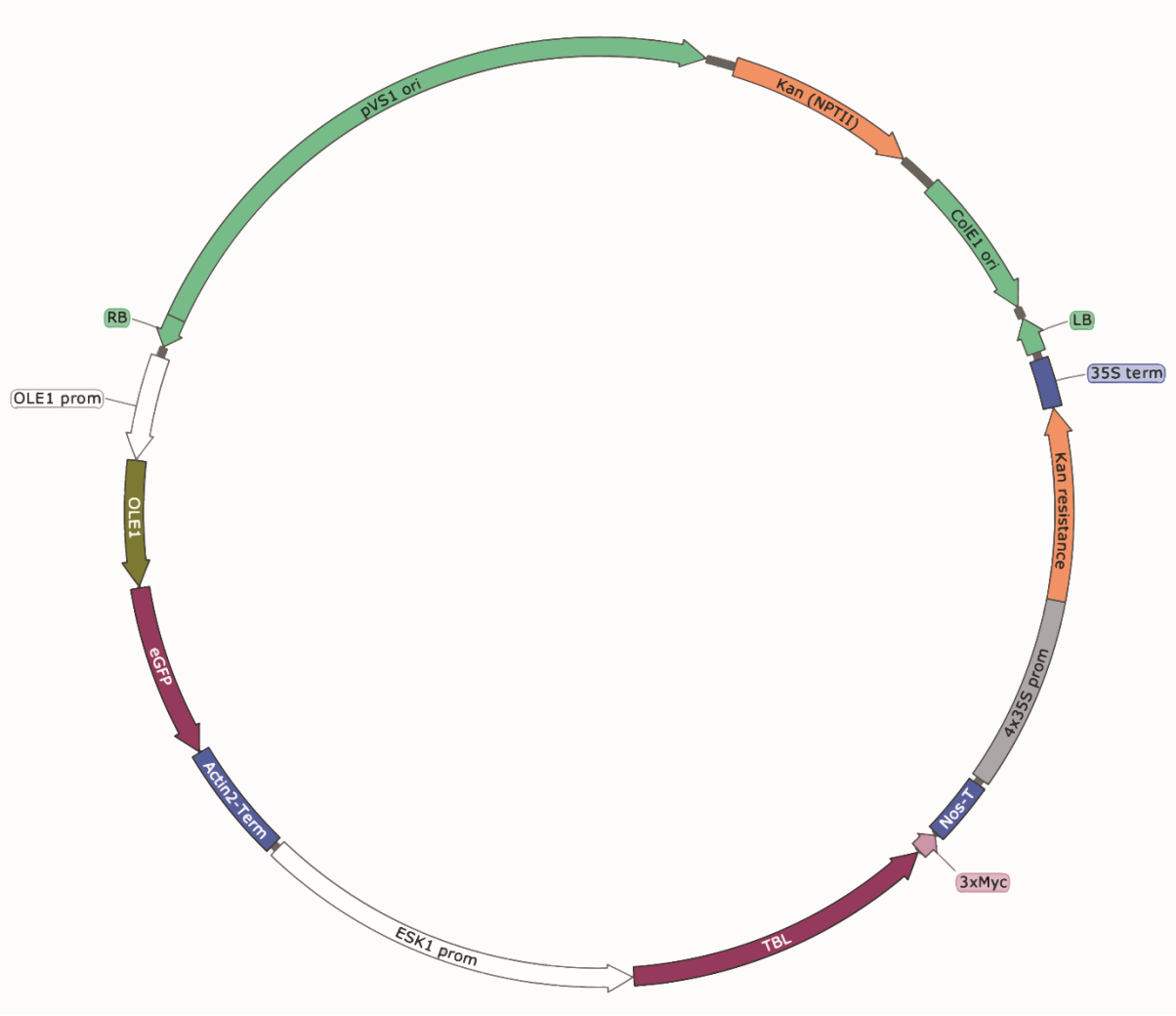


Figure 6.6: Vector for expression of *TBLs* under the promoter of *ESK1*.
These vectors were assembled by using Golden-Gate technology. The TBL coding regions were synthesised, verified by sequencing and cloned into the vector. The vector sizes vary between 12 Kb and 14 Kb. *Kan (NPTII)* is a bacterial kanamycin/neomycin resistance selection marker. *ColE1 ori* and *pVS1 ori* are bacterial origins for replication. *LB* and *RB* are flanking sequences for insertion in plants. *OLE1prom:OLE1:eGFP* is Oleosin-GFP fluorescent selection marker expressed in seeds. *TBL* coding regions used for independent assemblies: *TBL3*, *TBL28*, *TBL29*, *TBL30*, *TBL31*, *TBL32*, *TBL33*, *TBL34*, *TBL35* and *TBL36*.

6.2.2.2 Growth restoration of promoter swapped lines

The inactivation of ESK1 causes a reduction in the plant growth, among other effects (Xin and Browse, 1998). Thus, a complete restoration of the WT growth is expected when reintroducing a functional version of *ESK1* with its native promoter, since is the same gene that has been inactivated in *esk1* mutant. A similar restoration of WT growth in *esk1* by the expression of another *TBL* gene under *ESK1* promoter would suggest that such TBL may have a similar catalytic activity to ESK1. To investigate this, the transformed lines were grown in parallel for two weeks in agar plates, and then they were transferred to soil and grown for six weeks.

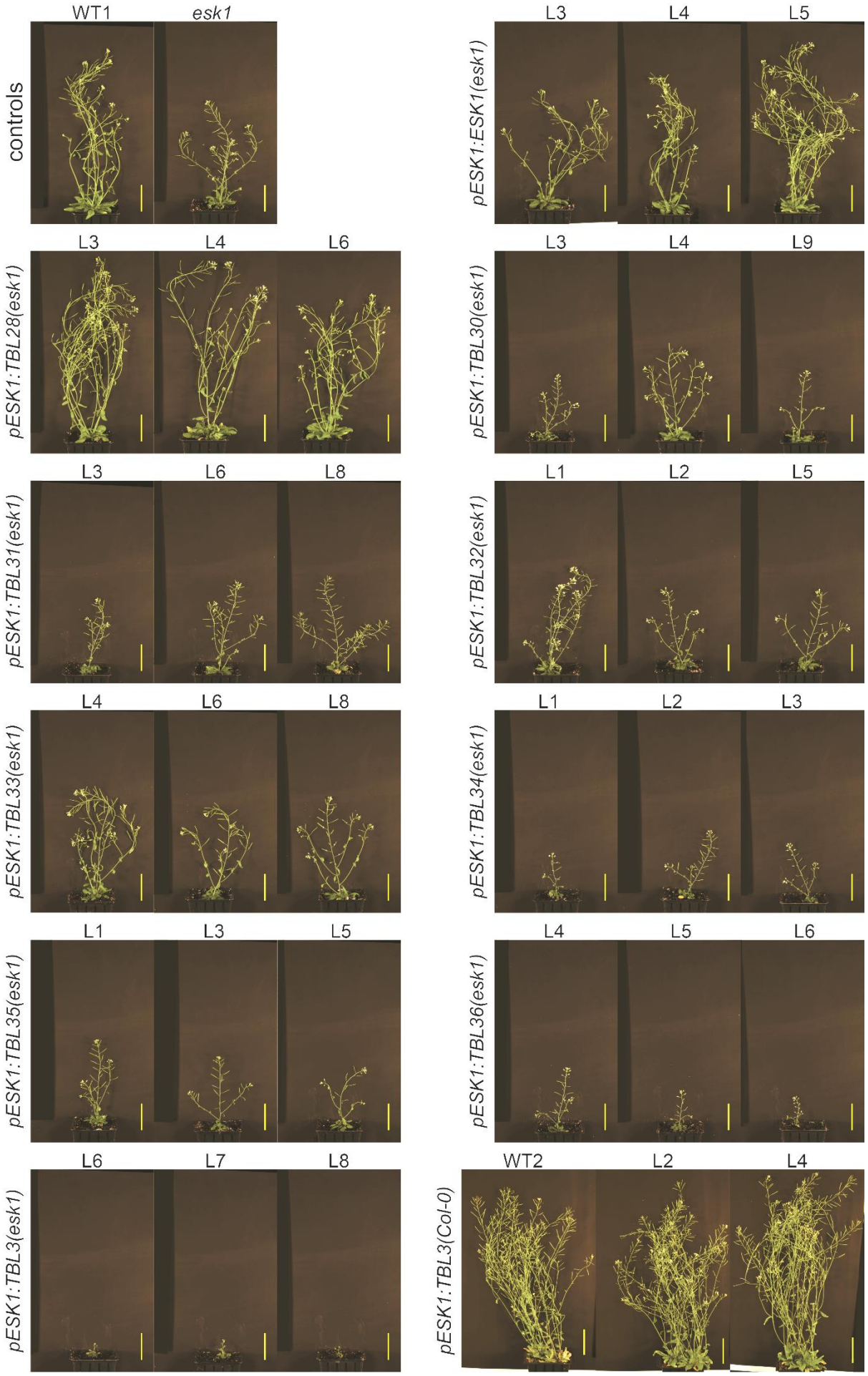
The growth of the promoter swapped plants is shown in Figure 6.7. As expected, the reduced growth of *esk1* was restored with the expression of *pESK1:ESK1(esk1)*, especially in L5. The *pESK1:TBL28(esk1)* lines had WT-like growth. This shows that TBL28 can provide the necessary function to allow growth restoration, strongly suggesting that the enzymatic activity of TBL28 may be similar to ESK1. This result shows that the activity of TBL28 in *esk1* mutant seems to be insufficient, but when boosted to the level of ESK1 with the *ESK1* promoter, growth is restored.

For the lines *pESK1:TBL30(esk1)*, *pESK1:TBL31(esk1)*, *pESK1:TBL32(esk1)*, *pESK1:TBL33(esk1)*, *pESK1:TBL34(esk1)*, *pESK1:TBL35(esk1)*, *pESK1:TBL36(esk1)* and *pESK1:TBL3(esk1)* there was no restoration of growth, suggesting that the respective TBL proteins may have different biochemical activities to ESK1. Dwarfing of *esk1* is variable and the height of these transformant lines were within the normal range. Surprisingly *pESK1:TBL3(esk1)* showed a more severe dwarfism in all the lines (height < 5 cm), including hemizygous lines for the insertion indicating a dominant effect (not shown). In addition, most *pESK1:TBL3(esk1)* lines did not produce seeds. The exacerbated dwarfism and infertility of *pESK1:TBL3(esk1)* lines suggesting that the expression of TBL3 in *esk1* produced an effect that is toxic for the plant growth.

Since the growth of *pESK1:TBL3(*Col-0) lines was similar to WT, TBL3 may have an effect on xylan, which makes it dysfunctional when the acetylation of xylan is reduced by the inactivation of ESK1. This result suggests that TBL3 may be active on xylan, in a different way to ESK1. It would be interesting to investigate further the xylan

acetylation of the *pESK1:TBL3(esk1)* lines, and verify the expression of TBL3 in the stem of the *pESK1:TBL3(*Col-0) lines.

Figure 6.7: Growth of promoter swapped plants.
esk1 and Col-0 lines were transformed with the vector shown in Figure 6.6. Representative plants are shown for each line. All lines except for *pESK1:TBL3(*Col-0*)* lines and WT1 were grown in parallel. *pESK1:TBL3(*Col-0*)* lines and WT2 were grown in parallel. Scale bar = 5 cm.


controls
WT1
esk1
pESK1:ESK1(esk1)
L3
L4
L5
pESK1:TBL28(esk1)
L3
L4
L6
pESK1:TBL30(esk1)
L3
L4
L9
pESK1:TBL31(esk1)
L3
L6
L8
pESK1:TBL32(esk1)
L1
L2
L5
pESK1:TBL33(esk1)
L4
L6
L8
pESK1:TBL34(esk1)
L1
L2
L3
pESK1:TBL35(esk1)
L1
L3
L5
pESK1:TBL36(esk1)
L4
L5
L6
pESK1:TBL3(esk1)
L6
L7
L8
pESK1:TBL3(Col-0)
WT2
L2
L4

6.2.2.3 Analysis of xylan GlcA pattern in promoter swapped lines

In the *esk1* mutant, the capability of xylan to bind to cellulose surface in a two-fold-screw is reduced, possibly because the [Me]GlcA pattern by GUX1 is altered (Grantham et al., 2017). The lack of this type of xylan-cellulose interactions is a possible cause of *esk1* dwarfed growth. The expression of two *TBL* genes in *esk1* restored the plant growth. *pESK1:ESK1* was expected to do so because it is the same protein that has been inactivated in *esk1*. However, the expression of *pESK1:TBL28* also showed a restoration of the plant growth, which suggests that TBL28 can provide the necessary function to allow growth restoration. Thus, restoration of the GUX1-mediated [Me]GlcA pattern of xylan is possible in *pESK1:ESK1(esk1)* and *pESK1:TBL28(esk1)* lines. In contrast, the expression of *pESK1:TBL3* showed an exacerbation of the dwarfism which is possibly caused by a further xylan loss of function. Differences in the substitution patterns on xylan may explain the differences observed in the restoration of growth by the overexpression of the *TBLs*. Since changes in acetylation in *esk1* affects [Me]GlcA patterns, the [Me]GlcA pattern of xylan in the promoter swapped lines was next investigated.

To analyse the [Me]GlcA pattern of xylan in the promoter swapped lines, the plants were grown for six weeks and basal stems were harvested from a pool of plants. AIR was prepared and xylan was extracted with alkali. The [Me]GlcA pattern was then analysed by PACE using GH30 xylanase (see Figure 6.8). In *esk1* the most abundant oligosaccharides released were $U^{[me]}X_6$, followed by $U^{[me]}X_7$, while in WT the main oligosaccharides were $U^{[me]}X_8$ and $U^{[me]}X_{10}$. As previously reported, GH30 products on *esk1* xylan showed no preference for xylooligosaccharides with even number of backbone Xyl, which is indicative of the altered GlcA pattern of xylan in this mutant (Grantham et al., 2017).

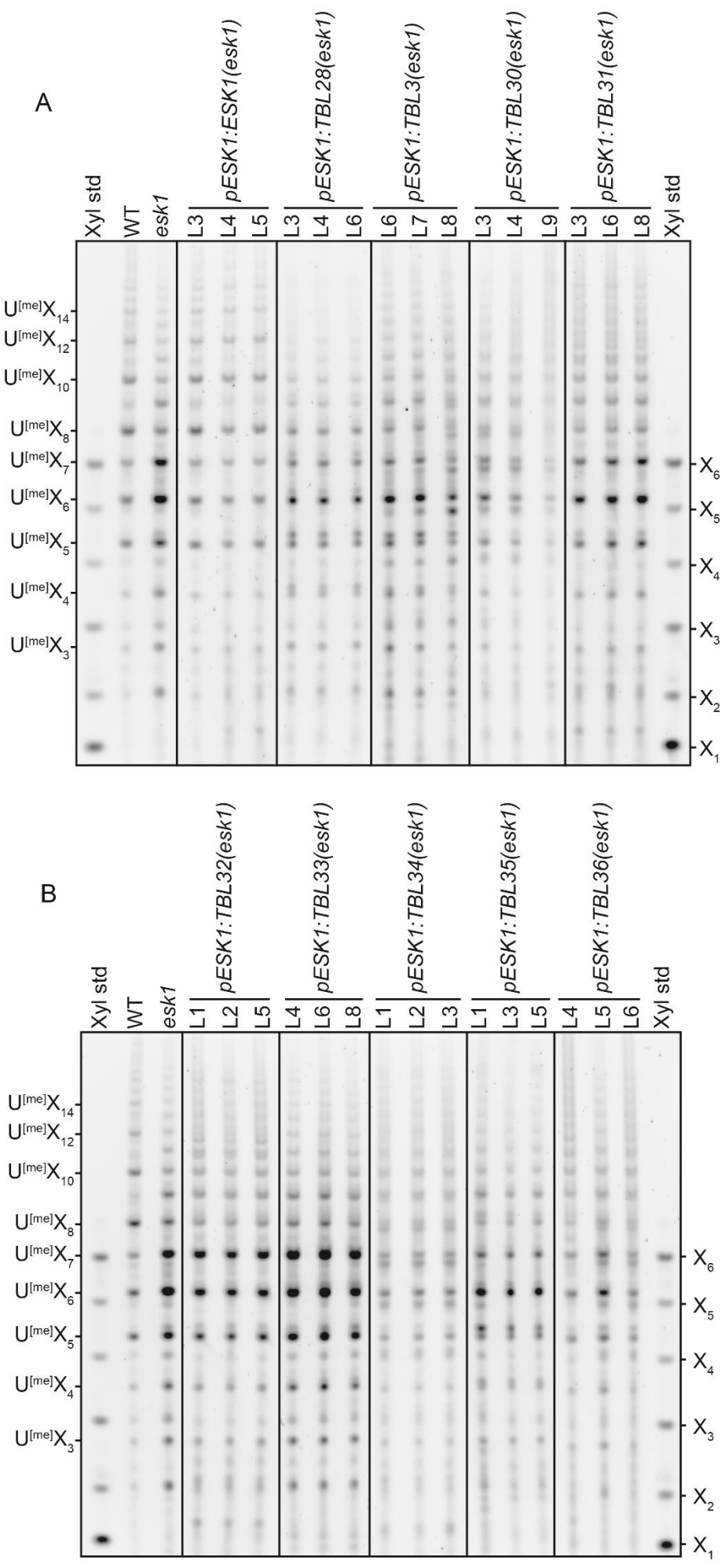
In the three *pESK1:ESK1(esk1)* lines (L3, L4 and L5), $U^{[me]}X_8$ and $U^{[me]}X_{10}$ showed a stronger band on PACE than $U^{[me]}X_7$, $U^{[me]}X_9$ and $U^{[me]}X_{11}$, indicating a higher frequency of even spacing of GlcA on xylan, and therefore indicating a restoration of the GUX1 GlcA pattern.

The *pESK1:TBL28(esk1)* and *pESK1:TBL3(esk1)* lines produced predominantly $U^{[me]}X_6$, which was similar to *esk1*, however, $U^{[me]}X_7$ was no longer a predominant GH30 product. The growth restoration and the phylogenetic relationship suggested that the activity might be similar, and therefore this result was unexpected. The proportion of $U^{[me]}X_8$ to $U^{[me]}X_7$ was increased in *pESK1:TBL28(esk1)* and *pESK1:TBL3(esk1)* lines compared to *esk1*, which suggests a partial restoration of the [Me]GlcA pattern.

The *pESK1:TBL32(esk1)* and *pESK1:TBL33(esk1)* lines did not show a restoration of the [Me]GlcA pattern of xylan. The ratio between $U^{[me]}X_6$ and $U^{[me]}X_7$ however, seems changed. In the transformed lines GH30 had a predominant production of $U^{[me]}X_7$ over $U^{[me]}X_6$, which is the opposite to the predominance exhibited by *esk1*.

The results for *pESK1:TBL30(esk1)*, *pESK1:TBL31(esk1)*, *pESK1:TBL34(esk1)*, *pESK1:TBL35(esk1)* and *pESK1:TBL36(esk1)* were unclear. Sometimes, GH30 xylanase miss-cuts the xylan producing artefact doublets, which have been observed in the Dupree lab. These doublets are produced when GH30 cuts in position +1 respect to the GlcA substitution, instead of cutting in +2. Despite the production of doublets, the mentioned transformants did not show a restoration of the [Me]GlcA pattern. Each line was analysed once, but at least three lines were analysed for each transformation. The [Me]GlcA pattern of xylan in the promoter swapped lines should be further analysed and perhaps quantitated to determine if there are any significant changes.

Figure 6.8 GlcA pattern of secondary-cell-wall xylan of the promoter swap transformants.
Plants with promoter swapped expression of *TBL* genes under the native ESK1 promoter in *esk1* background (*pESK1:TBL(esk1)*). T3 homozygous lines for the insertion were grown for six weeks, basal mature stems were collected, and AIR was prepared. Xylan was extracted with alkali and digested with BoGH30 glucuronoxylanase. The products were labelled with ANTS and resolved on a 10% PACE gel. WT and *esk1* controls are shown. Panels A and B were assembled from several gels.


A
Xyl std
WT
esk1
L3
L4
L5
pESK1:ESK1(esk1)
L3
L4
L6
pESK1:TBL28(esk1)
L6
L7
L8
pESK1:TBL3(esk1)
L3
L4
L9
pESK1:TBL30(esk1)
L3
L6
L8
pESK1:TBL31(esk1)
Xyl std
U[me]X14
U[me]X12
U[me]X10
U[me]X8
U[me]X7
U[me]X6
U[me]X5
U[me]X4
U[me]X3
X6
X5
X4
X3
X2
X1
B
Xyl std
WT
esk1
L1
L2
L5
pESK1:TBL32(esk1)
L4
L6
L8
pESK1:TBL33(esk1)
L1
L2
L3
pESK1:TBL34(esk1)
L1
L3
L5
pESK1:TBL35(esk1)
L4
L5
L6
pESK1:TBL36(esk1)
Xyl std
U[me]X14
U[me]X12
U[me]X10
U[me]X8
U[me]X7
U[me]X6
U[me]X5
U[me]X4
U[me]X3
X6
X5
X4
X3
X2
X1

6.2.3 The inactivation of *TBL* genes alters the xylan sidechain substitutions

ESK1 is not the only acetyltransferase that acts on xylan, because the xylan of the *esk1* mutant, where ESK1 has been inactivated, still carries Ac (Xiong et al., 2013). The promoter swap expression of the *TBLs* presented in the previous section, suggested that TBL28 may have a different enzymatic activity to ESK1 because it did not completely restore the [Me]GlcA pattern of xylan. Also, *esk1* showed an unexpected exacerbation of the dwarfism when transformed with *pESK1:TBL3*. When *esk1* was transformed with *pESK1:TBL32* and *pESK1:TBL33* the GlcA pattern of xylan was affected (but not partially restored). This evidence suggests that TBL3, TBL32 and TBL33 may have a different enzymatic activity and function to ESK1. The different Ac positions and patterns on *Arabidopsis* xylan (Chapter 4; Chapter 5; Teleman et al., 2000), and the conservation of ESK1 homologues in eudicots (this Chapter, section 6.2.1.2) support the hypothesis that multiple xylan-acetyltransferase activities are required for synthesis.

In the work described in this section, to investigate xylan-acetyltransferase activities additional to ESK1, plants with T-DNA insertions in the *TBL* genes were analysed, and the lack of expression of the respective *TBL* genes in the insertional mutants was verified by RT-PCR.

The effects on xylan caused by the inactivation of the *TBLs* were then analysed. Alterations in the acetylation and [Me]GlcA patterns of xylan in the *tbl* mutants were screened by PACE using combinations of GH11 xylanase, CE4 acetylesterase, and alkali treatment after hydrolysis.

6.2.3.1 Selection and isolation of stable *tbl* knockouts

To obtain knockout (KO) plants, collections of T-DNA insertional *Arabidopsis* lines have been made and stored in seedbanks such as SALK Institute or the Nottingham Arabidopsis Stock Centre. Lines carrying a single insertion in some genes are annotated and publicly available. Some of the insertions disrupt genes of interest and produce KO plants. Possible TBL insertional lines were obtained and the insertion positions were confirmed by PCR. Where possible, at least two independent insertion mutant alleles were studied. The insertional lines were then checked for the lack of

expression of the respective *TBL* genes by non-quantitative RT-PCR. The expression analysis of the *tbl33-1* mutant is shown in Figure 6.9. A similar outcome was produced in the expression analysis in leaf and stem of the other *tbl* mutants (not shown), and the result of this analysis is listed in Table 6.1. The plant lines analysed, and primers designed for their analysis are also listed in Table 6.1. The plants that had a full KO effect from the T-DNA insertion were selected to continue this work. The lines that had no insertion, or that the insertion produced partial or no reduction in the expression were discarded. The *tbl31-1* mutant was obtained towards the end of this project and no investigation was carried out with it. However, it would be interesting in future projects to investigate whether the inactivation of TBL31 has an effect on xylan biosynthesis.

Table 6.1: List of T-DNA insertional lines obtained and analysed.
PCR and RT-PCR primers used for the analysis of the respective mutants are listed. Insertion site was confirmed by sequencing (Seq results); correct insertions (OK) or other outcomes are shown. * *axy9* was described by Schultink et al. (2015). ** *esk1* homozygous mutant was provided by Mylene Durand-Tardif, and has been previously described in several reports (Grantham, 2016; Grantham et al., 2017; Xiong et al., 2013; Yuan et al., 2013).

									Expression in leaf		Expression in stem	
Gene	Locus	Allele	Mutant name	Lab code	Insertion site	PCR primers (RP/LP/BP)	Insertion sequence	RT-PCR primers	WT	Mutant	WT	Mutant
TBL28	AT2G40150	tbl28-1	SALK_025118	BHN	exon 3	1313/1332/248	OK	1564/1576	yes	no	yes	no
TBL28	AT2G40150	tbl28-2	SAIL_1215_F04	BHO	exon 2	1314/1333/405	OK	1564/1576	yes	no	yes	no
TBL3	AT5G01360	tbl3-1	SALK_149572	BHP	intron 3	1315/1334/248	misannotated					
TBL3	AT5G01360	tbl3-2	SALK_103316	BHQ	exon2	1316/1335/248	OK	1299/1334	yes	no	yes	no
TBL3	AT5G01360	tbl3-3	SALK_065969	BHR	exon 3	1317/1298/248	no insertion					
TBL3	AT5G01360	tbl3-4	SALK_066727	BHS	exon 2	1318/1299/248	no insertion					
TBL3	AT5G01360	tbl3-5	SALK_065959	BIF	exon 3	1381/1382/248	OK	1299/1334	no	no	yes	yes
TBL30	AT2G40160	tbl30-1	SALK_125287	BHT	exon 3	1319/1330/248	OK	1569/1581	yes	no	no	no
TBL30	AT2G40160	tbl30-2	SAIL_374_B10	BHU	exon 5	1320/1301/405	OK	1569/1581	no	no	no	no
TBL30	AT2G40160	tbl30-3	SALK_070679	BHV	exon 1	1321/1302/248	no amplification					
TBL30	AT2G40160	tbl30-4	SALK_070260	BHW	promoter	1322/1303/248	OK	1569/1581	no	no	no	no
TBL30	AT2G40160	tbl30-5	SALK_070221	BHX	promoter	1323/1304/248	no amplification					
TBL31	AT1G73140	tbl31-1	SALK_202988	BIO	intron	1623/1624/248	OK	1636/1623	no	no	yes	no
TBL32	AT3G11030	tbl32-1	SALK_125641	BHY	exon 5	1324/1305/248	OK	1305/1604	yes	no	yes	no
TBL32	AT3G11030	tbl32-2	SALK_074936	BHZ	exon 5	1324/1305/248	Identical to tbl32-1					
TBL33	AT2G40320	tbl33-1	SALK_015056	BIA	exon 3	1384/1383/249	OK	1327/1585	no	no	yes	no
TBL33	AT2G40320	tbl33-2	GK-548F05	BIB	intron 5	1328/1309/476	OK	1327/1585	no	no	yes	no
TBL34	AT2G38320	tbl34-1	GK-897H07	BIG	exon 1	1590/1591/476	OK	1573/1606	yes	yes	yes	yes
TBL34	AT2G38320	tbl34-2	SAIL_575_E10	BIK	promoter	1630/1631/405	OK	1632/1633	yes	yes	yes	yes
TBL34	AT2G38320	tbl34-5	RIKEN 11-2377-1	BIN	promoter	1573/1625/396	OK	1632/1633	yes	yes	yes	yes
TBL34	AT2G38320	tbl34-6	GABI_646H10	BIP	exon 3	1630/1635/210	OK	1632/1633	yes	yes	yes	yes
TBL34	AT2G38320	tbl34-3	SALK_006449.51.65	BIL	promoter	1631/1630/248	no insertion					
TBL35	AT5G01620	tbl35-1	SALK_007072	BIC	intron 5	1329/1310/248	OK	1338/1605	yes	no	yes	no
TBL35	AT5G01620	tbl35-2	GT_3_2546	BID	exon 1	1349/1336/1359	OK	1575/1330	yes	no	yes	no
TBL36	AT3G54260	tbl36-1	SALK_138715	BIE	exon 5	1331/1312/248	OK	1331/1312	no	no	yes	yes
AXY9*	AT3G03210	axy9-2	SALK_090612	BMX	exon 1	1629/1628/747						
ESK1**	AT3G55990	esk1-5	SALK_078275	BKD	exon 2	1234/1235/248						

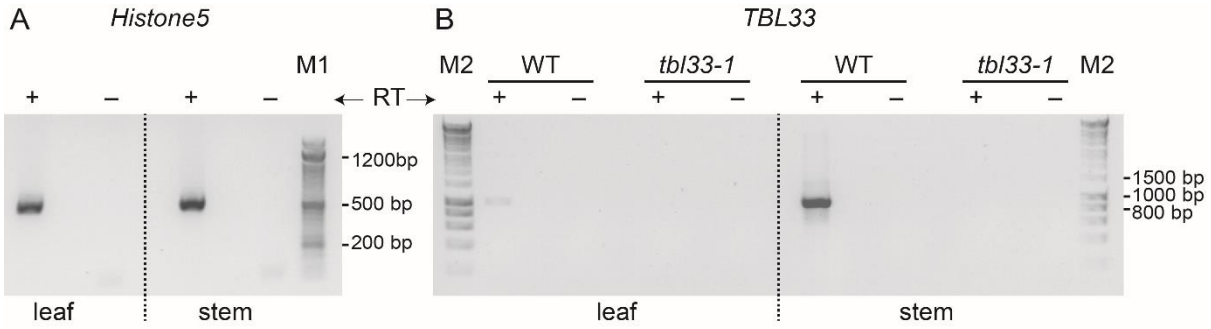


Figure 6.9: Expression analysis of *TBL33* in WT and *tbl33-1* mutant by RT-PCR. Leaf and stem tissues were harvested from 6-week-old plants, and RNA was extracted with a kit (Spectrum, Sigma-Aldrich), following the manufacturers' instructions. cDNA was produced from the template RNA using SuperScript™ III reverse transcriptase (Thermo fisher), reactions without retrotranscriptase controls (– RT) were used to identify genomic DNA contamination. The WT and *tbl33* cDNA samples were used for the PCR amplification of the gene *TBL33* (panel B), and of the control for DNA quality and quantity *Histone5* (panel A). The PCR of *Histone5* was carried out using primers 510 and 511, and the expected size was of 519 bp, and the PCR of *TBL33* was carried out with primers 1327 and 1585, and the expected amplification molecular size was of 989 bp (for primer sequences see Table 2.6 in Chapter 2, page 65). The PCR products were resolved on a 1.25% agarose gel in TAE buffer. The DNA molecular weight markers used were M1 (50bp DNA Ladder RTU) and M2 (HyperLadder 1kb, Bioline).

Some *TBL* genes are possibly redundant (this chapter, section 6.2.1.2), and therefore, the knockout of some might not show loss of function (if any). To anticipate this possible issue, double mutants *tbl28-1 esk1*, *tbl28-2 esk1*, *tbl32-1 tbl33-1* and *tbl32-1 tbl33-2* were produced. Additionally, because *esk1* has phenotypes on plant growth and xylan substitutions, crosses of *esk1* with other *tbls* could generate stronger phenotypes. In this way, if the lack of a TBL would not generate a clear effect, this effect could be exacerbated by the additional lack of ESK1. For this, crosses between *esk1* and the rest of the studied *tbls* were made (listed in Table 2.5, Chapter 2), and homozygous plants for both inactive genes were selected.

6.2.3.2 Growth of single T-DNA insertional lines and crosses

Some xylan-deficient mutants have growth phenotypes (Brown et al., 2009; Wu et al., 2010; Xiong et al., 2013), therefore, the easiest way to detect if the xylan is defective in the insertional mutants and crosses, is by observing the growth of the plant. Once the KOs were obtained and the lack of expression was confirmed, they were grown simultaneously to have the most accurate comparison. The plants were firstly grown

in sterile conditions on petri dishes for two weeks, then transferred to soil, and grown for four additional weeks in a long day regime (Figure 6.10).

All *tbl* single mutants showed a WT-like growth, except for *esk1*. Although the *tbl33-2* and *tbl35-2* mutants sometimes looked smaller than the WT this was not reproducible. The double mutants of the knockouts for putatively redundant genes *ESK1* and *TBL28* did not show a further effect on the growth to that caused by the mutation in *ESK1*. The double mutant of the knockouts for the putative gene duplicate pair *TBL33* and *TBL32* grew like WT. However, *esk1 tbl33-1* and *esk1 tbl33-2* showed a severe dwarfed growth, and neither produced seeds. Therefore, the inactivation of TBL33 and ESK1 may have an exacerbated effect on xylan, which may cause xylan to lose its function, affecting the growth of the plant. All other double mutants' heights were within the normal variation range of *esk1* mutant.

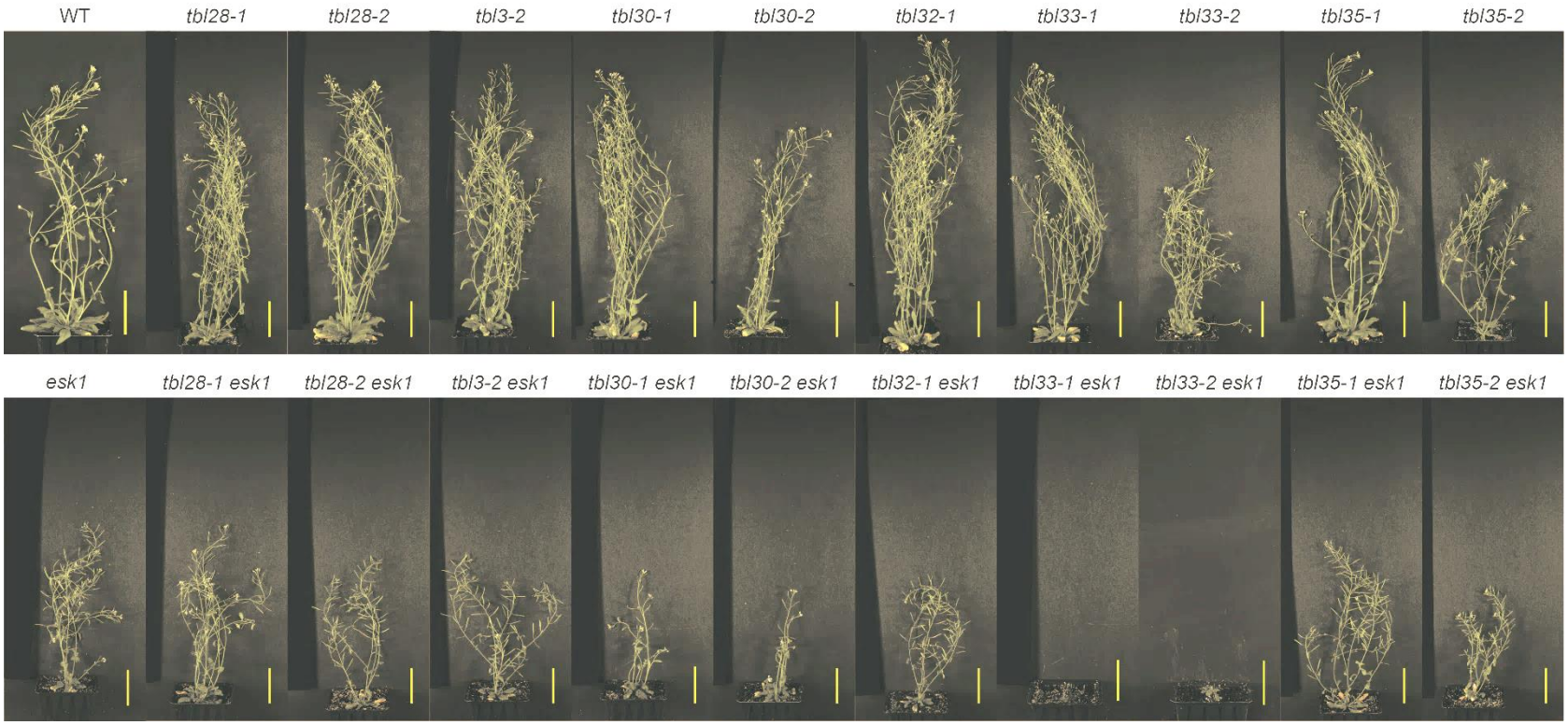


Figure 6.10: Growth of *tbl* T-DNA insertional lines and crosses between them.
In the top panel is shown the growth of WT and of all newly obtained lines containing a single T-DNA insertion that disrupts the expression of the respective gene. In the middle panel is shown the growth of the previously reported *esk1* mutant (*tbl29*) and of the crosses between *esk1* and other *tbl* mutants. In the bottom panel the crosses between the lines *tbl32* and *tbl33* are shown. All plant lines were grown in parallel. Scale bar = 5 cm.

6.2.3.3 Analysis of the GlcA substitution pattern of xylan in the *tbl* and *axy9* mutants

The inactivation of ESK1 causes a reduction of xylan acetylation, and also alters the [Me]GlcA pattern of xylan (Grantham, 2016). It is therefore possible that the inactivation of other acetyltransferase activities may cause an acetylation reduction, which may impact other downstream processes in the xylan biosynthesis pathway. AXY9 has recently been proposed to be part of the xylan acetylation process (Schultink et al., 2015), and its role has been hypothesised to stand close to the transfer of Ac to xylan by ESK1 and other TBLs. Therefore, the inactivation of AXY9 also may cause alterations of other xylan sidechains, by depleting downstream TBLs of acetylated substrate, and because of this the *axy9* mutant is also worth investigating.

To detect differences in the xylan sidechains of the *tbl* and *axy9* mutants, three detection methods were used. These methods rely on complete enzymatic digestions and separation of the produced oligosaccharides by PACE. The first approach consists of studying the frequency of the distances between the GlcA on the backbone. This was used previously to study the effect of the xylan glucuronyltransferases GUX1, GUX2 and GUX3 (Bromley et al., 2013; Mortimer et al., 2015). It was also used to show the effect of the lack of ESK1 on the xylan GlcA pattern (Grantham et al., 2017). The other two methods were developed here for directly detecting deficits in acetate groups on xylan, and are presented in sections 6.2.3.4 and 6.2.3.5.

The secondary-cell-wall xylan GlcA spacing analysis was carried out following the procedure outlined above (section 6.2.2.3), and is shown in Figure 6.11. As seen before, GlcA on WT secondary-cell-wall xylan are more frequently distanced by even number of Xyl (earlier in this Chapter; Chapter 4; Bromley et al., 2013). This pattern is revealed by the higher intensity of the species with even number of Xyl ($U^{[me]}X_6$, $U^{[me]}X_8$, $U^{[me]}X_{10}$) than odd ($U^{[me]}X_7$, $U^{[me]}X_9$, $U^{[me]}X_{11}$). In *esk1* the pattern is disrupted and the distances between the GlcA are smaller, therefore, the main products of GH30 are $U^{[me]}X_6$, and $U^{[me]}X_7$. Interestingly, *axy9* also showed an altered [Me]GlcA pattern of xylan. Because $U^{[me]}X_8$ is not the predominant GH30 product, it is clear that the lack of AXY9 has an effect on the [Me]GlcA pattern. However, *axy9* produced more $U^{[me]}X_8$

than $U^{[me]}X_7$, which implies that the alteration of the xylan [Me]GlcA pattern in *axy9* is not as severe as in *esk1*. In contrast to *esk1*, none of the other *tbl* mutants analysed show clear changes in the xylan [Me]GlcA pattern. There are several possible reasons why this pattern may be unaffected. Perhaps, the acetylation carried out by the other TBLs is not related to the patterning of the GlcA in the synthesis. It is also possible that their function may be related to primary-cell-wall xylan synthesis, which is not analysed in this experiment. Alternatively, some of the genes could be redundant. ESK1 however, is the only TBL analysed that has a disruption of the [Me]GlcA pattern of secondary-cell-wall xylan, and the partial loss of the pattern exhibited by *axy9* may be caused only by the depletion of substrate for ESK1 in this mutant.

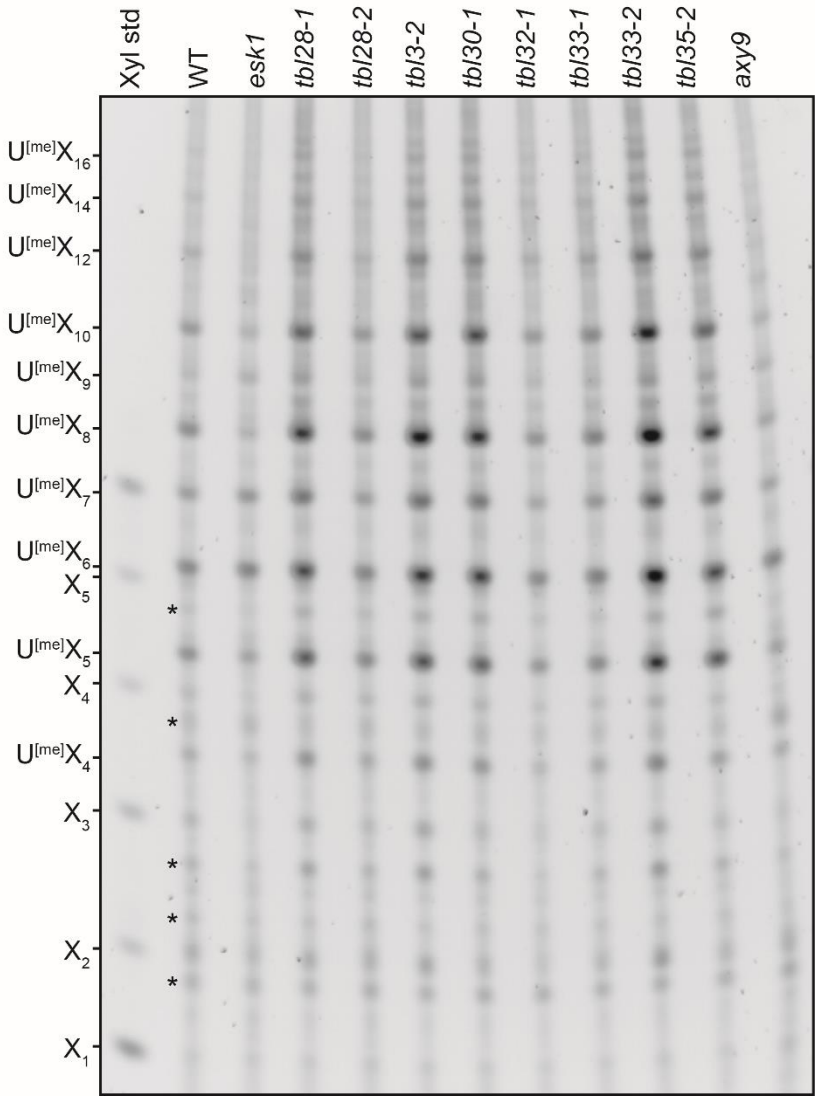


Figure 6.11 GlcA pattern of secondary-cell-wall xylan in the *tbl* mutants and *axy9*. AIR was prepared from mature bottom stems from the *tbl* mutants, and from whole stems from axy9 mutant. Xylan was extracted with alkali and hydrolysed with BoGH30 xylanase. The products were labelled with ANTS and resolved on a 10% PACE gel. * unidentified bands.

In Chapter 5 it was shown that stem and callus primary-cell-wall xylans are acetylated. These xylans are decorated with a [Me]GlcA every sixth Xyl, and some of these GlcA may be substituted with an Ara*p* (PUX) (Chapter 5; Mortimer et al., 2015; Peña et al., 2016). It is unknown if acetylation of xylan in primary walls determines the spacing pattern of the [Me]GlcA, in an analogous manner to the acetylation by ESK1 on secondary-cell-wall xylan. It is also unknown which TBLs are involved in the acetylation of xylan in primary cell walls.

Young stem cell walls are enriched for primary cell walls, compared to the bottom portion of the stem (Mortimer et al., 2015). Because the xylan GlcA pattern in this material has been previously described, and because of its ease of preparation, young stems were used for the analysis of primary-wall xylan in the *tbl* mutants.

The 6-Xyl spacing between [Me]GlcA substitutions on primary-cell-wall xylan from young stems was revealed by a characteristic doublet on PACE by GH30 products (Chapter 5; Mortimer et al., 2015). This doublet corresponds to the oligosaccharides $U^{[me]}X_6$ and PUX_6. On the other hand, the products with higher DP reveal the inevitable contamination from secondary cell walls in the young stem samples. These higher DP species only produce singlets, because PUX sidechains are only present in primary-wall xylan. If GUX3 was dysregulated analogously to GUX1 in secondary walls of *esk1* (Grantham et al., 2017), the spacing between PUX sidechains on xylan would vary, and therefore, GH30 products would produce doublets on PACE for the species with higher DP (e.g. $U^{[me]}X_7$ and PUX_7, $U^{[me]}X_8$ and PUX_8, etc).

To investigate the [Me]GlcA substitution pattern of primary-wall xylan in the *tbl* mutants, young stems were collected and AIR was prepared as described by Mortimer et al. (2015). Then the xylan was extracted with alkali and hydrolysed with GH30 glucuronoxylanase. The PACE analysis of the GH30 products is shown in Figure 6.12. As previously described, GH30 products on the WT sample produced a doublet corresponding to $U^{[me]}X_6$ and PUX_6, and also higher DP bands which are coherent with the GlcA pattern of secondary-cell-wall xylan (Bromley et al., 2013; Mortimer et al., 2015). *esk1* shows its characteristic altered pattern in the higher DP bands, but interestingly, the doublet of $U^{[me]}X_6$ and PUX_6 was produced without alteration. Despite the dysregulation of GUX1 and GUX2 in *esk1* mutant, the production of the doublet

only for $U^{[me]}X_6$ and PUX_6 and not for higher DP bands suggests that GUX3 activity is not dysregulated.

None of the *tbl* mutants showed a clear alteration of the [Me]GlcA pattern of primary-cell-wall xylan. As discussed above, some *TBL* genes may be redundant, and therefore, it would be interesting to investigate whether the inactivation of multiple TBLs have an effect on the [Me]GlcA substitutions on primary-cell-wall xylan. Perhaps TBL31 and TBL34 could have an analogous role to ESK1 in primary cell walls, but insertional mutants for *TBL31* and *TBL34* were unavailable at the time. To test whether some of the TBLs are redundant it would be interesting to study young stem material of *tbl tbl* double mutants, and include *tbl31* and *tbl34* in the analyses.

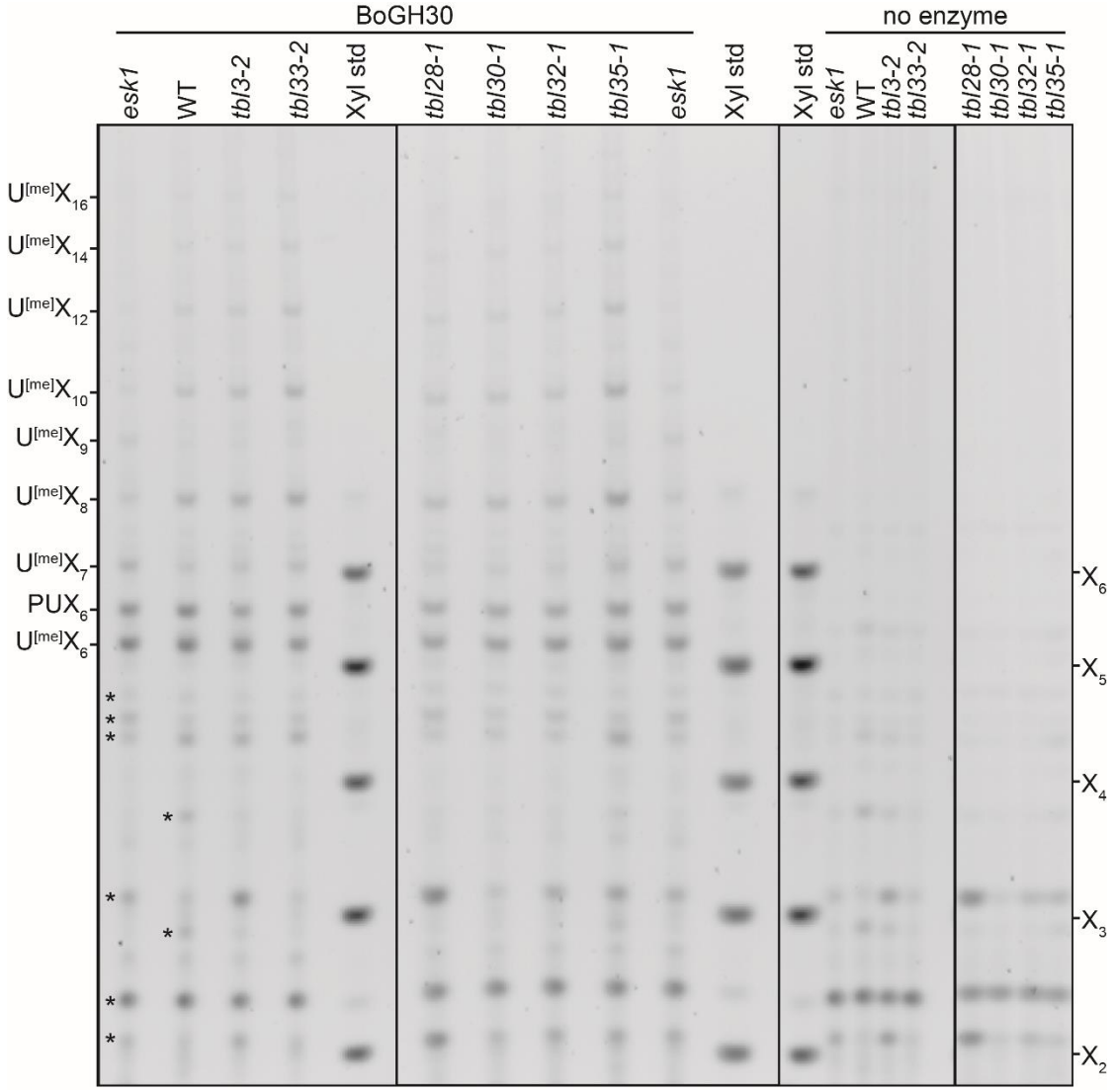


Figure 6.12 GlcA pattern of primary-cell-wall xylan in the *tbl* mutants
AIR was prepared from young stems, treated with alkali and hydrolysed with BoGH30 xylanase. The products were labelled with ANTS and resolved on a 10% PACE gel. * background bands. This figure is assembled from two gels.

6.2.3.4 Differences in the main alternating pattern of acetylation in the *tbl* mutants

ESK1 was the only TBL analysed whose activity determines the downstream xylan glucuronidation process in secondary walls (section 6.2.3.3; Grantham et al., 2017). Because the phylogenetic analysis in section 6.2.1 suggested that the TBLs from the ESK1 clade may have different activities on xylan, it is possible that the acetylation of xylan by other TBLs may not have a role in determining the pattern of [Me]GlcA. Therefore, it is possible that some *tbl* knockouts may have a deficit in xylan acetylation without having affected the [Me]GlcA pattern. To detect possible deficits in xylan acetylation in the *tbl* mutants two methodologies were developed and implemented. In this section the detection of Ac deficits in the main alternating acetylation pattern of xylan in the *tbl* mutants is presented, and in the next section a method for the detection of Ac associated with di-substituted backbone Xyl is shown.

In *Arabidopsis* xylan, the main acetylation pattern has been described as acetylated Xyl alternated with undecorated Xyl (Chapter 4; Busse-Wicher et al., 2014). If an Ac in this pattern is absent, a gap of three consecutive undecorated residues on the backbone is produced. Xylan with the complete alternating acetylation pattern is theoretically inaccessible by GH11 xylanase, but when there is gap in the Ac pattern, GH11 should be able to access the backbone in this gap and hydrolyse it (for GH11 activity see section 1.5.2 in Chapter 1, page 43,). Because *esk1* has been described to have a xylan acetylation deficit (Xiong et al., 2013), its xylan should be more accessible by GH11 than WT. To detect whether the inactivation of the TBLs produces a gap in the alternating acetylation pattern, the sensitivity of the xylan from the *tbl* mutants to GH11 was tested. In order to do this, AIR was prepared from bottom stems. To allow GH11 to access the xylan, the AIR was first depectinated with 5% ammonium oxalate and delignified with 11% peracetic acid. The resulting acetylated material (holocellulose) was incubated with GH11. The enzyme was inactivated with heat and removed using Nanosep. Because migrating Ac on oligosaccharides produce smeary bands on PACE (Busse-Wicher et al., 2014), to increase the resolution the hydrolysis products were treated with alkali. Then the oligosaccharides were labelled with ANTS and analysed on PACE (see Figure 6.13).

WT xylan produced a small amount of X_1, X_2, X_3, X_4 and $XU^{[me]}XX$, which indicates that GH11 had a certain level of accessibility to the xylan, possibly due to naturally occurring Ac gaps or a small Ac loss during the material preparation. The low amount of X_1 and X_2 indicates that the gaps in the acetylation pattern that occur in WT are not often consecutive gaps. Since the GH11 hydrolysis was complete (control not shown), X_3 and X_4 are produced only from acetylated species, indicating naturally occurring gaps in the alternated pattern. This also indicates that GH11 minimal acetylated products have three or four backbone residues.

GH11 hydrolysis of *esk1* holocellulose produced higher amount of product. This is because the acetylation in *esk1* xylan is decreased (Xiong et al., 2013). The higher production of X_1, and X_2 indicates that GH11 had further access to the backbone, which may be caused by consecutive gaps on the alternating acetylation pattern. The higher production of $XU^{[me]}XX$ indicates that the accessibility of GH11 was increased near the [Me]GlcA substitutions. The increased production of X_3 suggests that GH11 had accessibility to more acetylated Xyl flanked by gaps in the alternating acetylation pattern. In contrast, the production of X_4 by GH11 was not increased in *esk1*, which suggests that GH11 activity distinguishes around various Ac contexts.

Other *tbl* mutants did not exhibit a clear production of additional products. This could mean that either the respective TBL proteins may not have a function associated with xylan, they may be redundant to each other, or that they may acetylate a domain where Acs are much closer together and a loss of Ac is less likely to open a gap in which GH11 can access. Alternatively, their activity may be specific for primary-cell-wall xylan which was not assessed here.

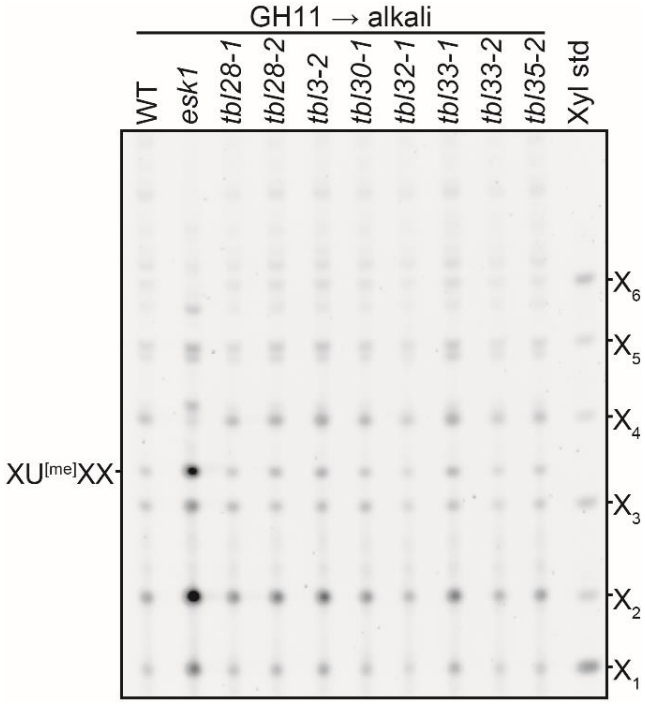


Figure 6.13 GH11 accessibility test on *tbl* mutants holocellulose.
Holocellulose was prepared from mature bottom stems, and digested with GH11 xylanase. The enzyme was inactivated with heat and removed using Nanosep. The hydrolysis products were treated with alkali, labelled with ANTS and resolved on a 10% PACE gel.

6.2.3.5 Analysis of Ac associated with di-substituted xylan residues

In the previous experiment, the main alternating acetylation pattern of secondary-cell-wall xylan was analysed in the *tbl* mutants by the detection of gaps. However, the TBLs may have activities associated to other regions of xylan. In addition to the regular pattern of xylan, Ac can co-substitute a Xyl that carries an additional substitution, such as Ac, an Ara*f* or a [Me]GlcA (Chapters 4 and 5).

To determine whether the inactivation of the TBLs affect Ac on xylan present in di-substituted Xyl, the xylan from the *tbl* mutants was analysed by PACE using CE4 acetylesterase in combination with GH11. The acetylesterase CE4 removes all single acetate groups from xylan, while $U^{[me]}$, $U^{[me],a}$ and X^{aa} are unaffected (Puchart and Biely, 2015). When CE4 and GH11 are used in combination, the predominant products are X, XX and XRXX (where R stands for all substituted xylan backbone residues unaffected by CE4). Also, depending on how close Rs are, some minor products can be produced, such as XRRXX, XRXRXX, XRXXRXX, XRXRXRXX, etc. The lack of Ac at different positions would produce a different effect on PACE when using CE4/GH11. If di-acetylated Xyl are affected by the inactivation of a TBL, then the production of $XX^{aa}XX$ would be affected. If an effect is produced on Xyl di-substituted with [me]GlcA and Ac, then the production of $XU^{[me],a}XX$ and/or $XU^{[me]}XX$ would be altered. Changes in Xyl mono-substituted with Ac would not be detected using this approach, and any effect on the production of longer oligosaccharides would suggest a deficit in Ac related to neighbouring Rs.

To digest the xylan, holocellulose from the bottom stems of the *tbl* mutants was co-incubated with GH11 and CE4. Then the enzymes were inactivated with heat and removed using Nanosep. The PACE analysis of the products is shown in Figure 6.14 A, and to differentiate between the species $XU^{[me],a}XX$ and $XU^{[me]}XX$, the GH11/CE4 digestion products were treated with alkali (panel B).

The WT xylan produced the expected oligosaccharides X_1, X_2, $XU^{[me],a}XX$, $XX^{aa}XX$ and some higher DP species (A–D). In panel B, the de-acetylated species correspond to X_1, X_2, $XU^{[me]}XX$, X_4, X_5, and some unknown species E–G. The hydrolysis of xylan of the mutants *esk1*, *tbl3*, *tbl33-1* and *tbl33-2* exhibited clear differences to the WT.

In *esk1* the species $XX^{aa}XX$ was undetectable, and after the alkali treatment the species X_4 was also absent. The absence of $XX^{aa}XX$ in GH11/CE4 products on *esk1* xylan indicates that in *esk1* the frequency of X^{aa} residues is substantially reduced and possibly completely absent. Supporting this, after the alkali treatment, *esk1* xylan did not produce X_5, which is also an oligosaccharide produced by GH11/CE4 from regions of xylan that are substituted only with Ac. Because the bands A, B and C were also absent in *esk1* products, these three oligosaccharides may contain at least one X^{aa} in their structure. On the other hand, the species D was produced by *esk1* xylan which suggests it has $U^{[me],[a]}$ types of residues in its structure and/or X^{aa} not mediated by ESK1. On panel B, *esk1* produced $XU^{[me]}XX$ and band G, and both increased compared to WT. $XU^{[me]}XX$ in *esk1* is produced by the de-acetylation of $XU^{[me],a}XX$, and band G is produced when D is de-acetylated. This is inferred by a similar abundance prior and after alkali treatment of *esk1* products. Furthermore, G also co-migrated with the characterised species $XU^{[me]}XU^{[me]}XX$ (Łyczakowski, 2018; not shown), which means that D could correspond to $XU^{[me],a}XU^{[me],a}XX$. In addition, the smear below band D (double asterisk) could be $XU^{[me],a}XU^{[me]}XX$ and/or $XU^{[me]}XU^{[me],a}XX$, which after the treatment with alkali would become G.

Both *tbl33-1* and *tbl33-2* produced an extra band that co-migrated with $XU^{[me]}XX$, which indicates that in these mutants, some (methyl)glucuronidated Xyl are found unacetylated. This implicates TBL33 in acetylation of this Xyl. However, some $XU^{[me],a}XX$ was also detected, which implies that TBL33 is only partially responsible for Acs on (methyl)glucuronidated Xyl. After treating with alkali, the doublet $XU^{[me],[a]}XX$ became only $XU^{[me]}XX$, shifting into a unique band for *tbl33*.

The species A and E were absent in *esk1*, *tbl33-1* and *tbl33-2*. Possibly because the product A may have X^{aa} in its structure (discussed above). In addition, the absence of A in *tbl33* suggests that A may have $U^{[me],a}$ in its structure, because in *tbl33* this site-specific Ac is partially missing for $XU^{[me],a}XX$.

Band B was not produced by *esk1* and was slightly more abundant in *tbl3*. In comparison, X_5 in panel B followed the same behaviour. This strongly suggests that B is $XX^{aa}X^{aa}XX$. This also suggests that the function of TBL3 could be to add Ac near $X^{aa}X^{aa}$, or to remove Ac from $X^{aa}X^{aa}$.

Band C was produced by all samples except for *esk1* and *tbl3*. Also, band C seemed slightly stronger in *tbl33-1* and *tbl33-2*, which suggests that it may have a non-acetylated $U^{[me]}$ residue. After the alkali treatment, the species F followed a similar pattern to band C, because it was not produced by *esk1* and *tbl3*, and slightly stronger in *tbl33-1* and *tbl33-2* relation to X_5. This suggests that F may be produced when C is treated with alkali, however, this behaviour of C and F did not show reproducibility in another mutant (section 6.2.4.3).

The results were reproducible, and a summary of the putative identities for the bands on the gel are shown in Table 6.2. The identities of the CE4/GH11 products were further analysed by MS in section 6.2.4.1.

None of the other mutants showed a clear difference to the WT. The results shown here strongly suggest that TBL33 and TBL3 are active on secondary-cell-wall xylan. It was also shown that ESK1 is necessary for the synthesis of di-acetylated Xyl on secondary-cell-wall xylan.

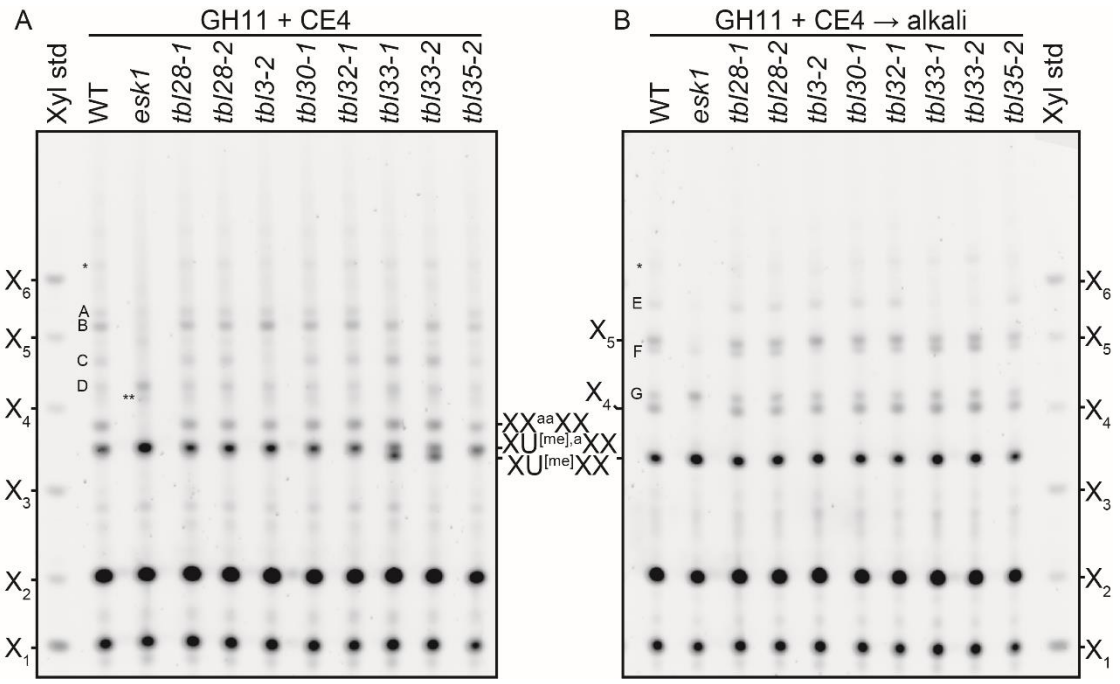


Figure 6.14 Detection of differences associated with di-substituted Xyl.
Holocellulose was prepared from mature bottom stems, and digested simultaneously with GH11 xylanase and CE4 acetylesterase (panel A). The enzymes were heat-inactivated and removed with Nanosep. The hydrolysis products were treated with alkali (panel B), labelled with ANTS and resolved on a 10% PACE gel. (*) contaminant band unaffected by alkali. (**) smear possibly corresponding to a mixture of $XU^{[me],\mathbf{a}}XU^{[me]}XX$ and $XU^{[me]}XU^{[me],\mathbf{a}}XX$. A-D are unidentified acetylated species. E-G are produced by acetylated species when treated with alkali. The putative identities of bands A-G are discussed in the text and a summary is presented in Table 6.2.

Table 6.2: Possible identities of CE4/GH11 product according to their behaviour.
The behaviour of the species is shown in Figure 6.1 and described in this section. The putative identities or shifts of CE4/GH11 products on holocellulose prepared from basal stems, are here listed. The identity of the bands marked with letters was not fully identified.

CE4/GH11 product			**After alkali treatment**	
A	X_6 Ac_3 $U^{[me]}$	→	**E**	X_6 $U^{[me]}$
B	X_5 Ac_4	→		X_5
C		→	**?**	
?		→	**F**	
D	$XU^{[me],a}XU^{[me],a}XX$	→	**G**	$XU^{[me]}XU^{[me]}XX$
******	$XU^{[me],a}XU^{[me]}XX$ and/or $XU^{[me]}XU^{[me],a}XX$	→	**G**	$XU^{[me]}XU^{[me]}XX$
	$XX^{aa}XX$	→		XXXX
	$XU^{[me],a}XX$	→		$XU^{[me]}XX$

6.2.4 Further studies on *tbls* and *axy9*

In the previous sections of this chapter, the objective was to identify possible acetyltransferases acting on xylan. A list of putative genes was made and a few of the candidates were discarded. The following genes shown to be active on xylan and their lack causes an alteration of the sidechains, and even in some cases, an effect on the growth: TBL3, TBL28, TBL29/ESK1 and TBL33. All of them were taken forward for additional analysis. TBL32 was also analysed further because it is a possible duplication of TBL33 (section 6.2.1.2). AXY9 has been reported to be involved in the xylan acetylation process (Schultink et al., 2015) and it was shown here to have an altered GlcA pattern, therefore it was also included in the next experiments.

6.2.4.1 TBL32 and TBL33

The *tbl33* mutant has a partial deficit in the Ac situated on (methyl)glucuronidated Xyl. Thus, another protein has a similar function and acetylates some of these sites. The best possible candidate is TBL32, because is the closest homologue to TBL33. If the respective genes are redundant, the double mutant (*tbl32/33*, section 6.2.3.2) should show a stronger phenotype. The first approach was to analyse the Ac associated with di-substituted Xyl in the double mutant xylan, similar to analysis explained in section 6.2.3.5. Holocellulose was prepared from *tbl32*, *tbl33* and *tbl32/33*. They were co-incubated with CE4 and GH11, and then treated with alkali. The analysis on PACE of

the products is shown in the Figure 6.15. In panel A the products are shown of the digestion with GH11 and CE4, and in panel B are shown the samples after alkali treatment. The unknown bands were labelled from A to G similar to Figure 6.14, and the possible identities are listed in Table 6.2. *tbl33* exhibited the doublet corresponding to $XU^{[me],a}XX$ and $XU^{[me]}XX$, which was observed previously in this chapter. Strikingly, *tbl32/33* produced only the unacetylated species. This shows that TBL32 is responsible for the acetylation of some of $U^{[me],a}$ sites, and that it has a similar function to TBL33. Remarkably, the additive effect of *tbl32* with *tbl33* in the double mutant made $XU^{[me],a}XX$ undetectable by this method. Both *tbl33* and *tbl32/33* did not produce species A and E. In *tbl33*, X_5 was missing but this was not reproducible. The band F seemed to be stronger in *tbl33* and *tbl32/33* compared to X_5. The allele *tbl33-2* showed a similar behaviour to *tbl33-1*, even when crossed with *tbl32-1* (not shown).

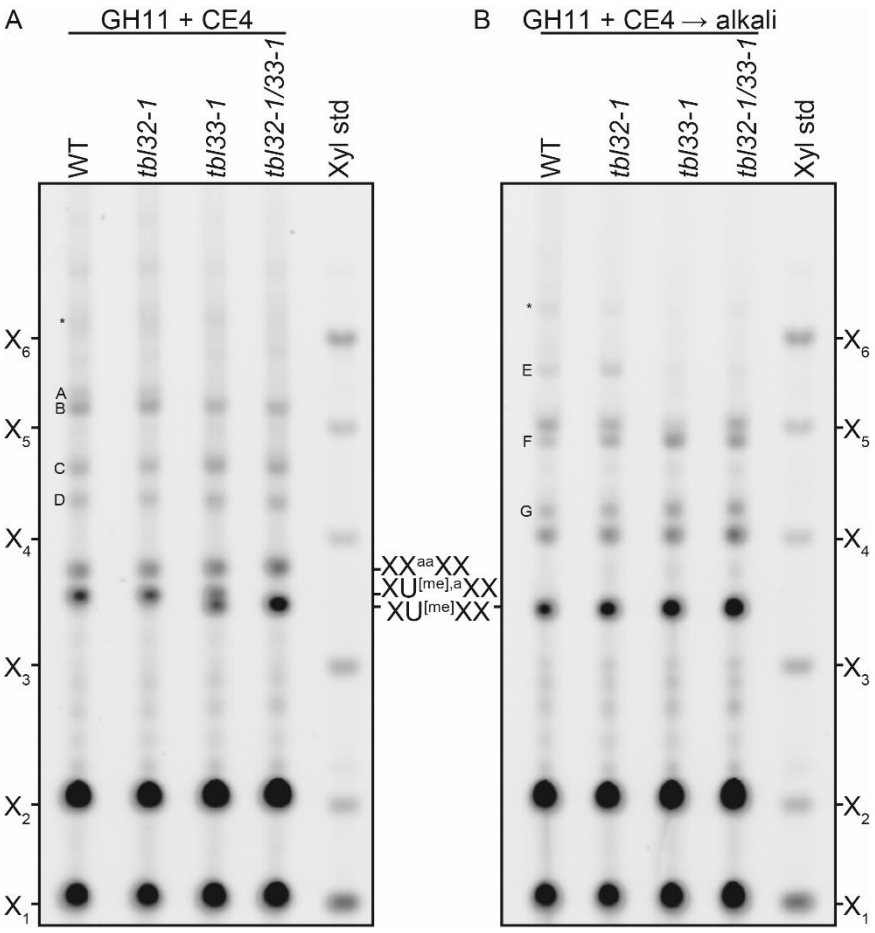


Figure 6.15: Double substitution on Xyl analysis of *tbl32/33*.
Holocellulose was prepared from mature bottom stems and digested simultaneously with GH11 and CE4 (panel A). The enzymes were heat-inactivated and removed using Nanosep. The hydrolysis products were treated with alkali (panel B), labelled with ANTS and resolved on a 10% PACE gel.

The PACE analyses shown above and in section 6.2.3.5 revealed that the CE4/GH11 products on *esk1* and on *tbl32/33* xylan are different. The deficiency of xylan acetylation exhibited by *tbl32/33* was different to *esk1* possibly because ESK1 and TBL32/33 are responsible for the transfer of Ac at different positions to the xylan. To confirm the identity of the CE4/GH11 products analysed on PACE, and to determine the position of the missing Ac in the mutants, GH11/CE4 products of WT, *esk1*, *tbl32*, *tbl33* and *tbl32/33* were analysed by mass-spectrometry. The unlabelled CE4/GH11 products were purified with a HyperSep hydrophobic interaction column, and then mixed with MS matrix DHB and spotted onto a MALDI plate. The MS analysis was carried out in a BRUKER MALDI mass-spectrometer in reflector mode, and the spectra obtained are shown in Figure 6.16.

The GH11/CE4 products were classified by the number and type of sidechains and the number of backbone residues (colour labels), and differences found in the mutants are next explained.

The species in brown arise from non-glucuronidated regions of xylan. *M/z* = 653 ($Pent_4Ac_2$) is present in WT and in the mutants *tbl32*, *tbl33* and *tbl32/33*, and is absent in *esk1*. This oligosaccharide on PACE ($XX^{aa}XX$) was a predominant product (after $XU^{[me],[a]}XX$) produced by WT, *tbl32*, *tbl33* and *tbl32/33*. Perhaps non-acidic oligosaccharides fly less efficiently than glucuronidated oligosaccharides in the mass-spectrometer. However, the absence of this peak in *esk1* spectrum is consistent with the analysis on PACE, in which $XX^{aa}XX$ was also undetectable.

The species in orange (with four backbone Pent) arise from [Me]GlcA substituted xylan Xyl that are isolated from other R (resistant to acetylesterase CE4). The isolated (methyl)glucuronidated Xyl detected were either acetylated or non-acetylated. WT, *tbl32* and *esk1* showed a preference for the species $Pent_4Ac_1U^{me}$ (*m/z* = 801) which is acetylated and methylated, followed by $Pent_4Ac_1U$ (*m/z* = 787) which is its unmethylated version. The signal height of $Pent_4Ac_1U$ (*m/z* = 787) was about half of $Pent_4Ac_1U^{me}$ (*m/z* = 801) for the WT, and a third for *esk1*, which is consistent with the degree of methylation of GlcA described before for WT and *esk1* (Grantham, 2016; Li et al., 2013; Peña et al., 2007). Instead, in *tbl33* the production of $Pent_4Ac_1U$ (*m/z* = 787) was on the noise levels, $Pent_4Ac_1U^{me}$ (*m/z* = 801) was also reduced, and the non-acetylated species $Pent_4U$ (*m/z* = 745) and $Pent_4U^{me}$ (*m/z* = 759) were increased. This

is indicative that in *tbl33* a larger proportion of [Me]GlcA-substituted Xyl are not acetylated. In the double mutant *tbl32/33* the production of the acetylated species (*m/z* = 787 and 801) was almost undetectable, instead, *m/z* = 745 and 759 were two of the four predominant peaks of the entire spectrum. This shows that in *tbl32/33* the (methyl)glucuronidated Xyl that are distant from other R-type substitutions, are not acetylated. The partial and complete loss of Ac in $Pent_4Ac_1U^{[me]}$ exhibited by the MS spectra of CE4/GH11 products on *tbl33* and *tbl32/33* xylan, is consistent with the mobility of $XU^{[me],a}XX$ on the PACE analysis above, and supports that the lack of Ac in *tbl32/33* is specifically at (methyl)glucuronidated Xyl.

The species labelled with red have one [Me]GlcA, and as they have six pentoses, they arise from regions of xylan where $U^{[me],[a]}$ and X^{aa} are at two Xyl of distance. Because of the activity of GH11, in WT *m/z* = 1149 is a mixture of $XU^{me,a}XX^{aa}XX$ and $XX^{aa}XU^{me,a}XX$, and *m/z* = 1135 is a mixture of $XU^{a}XX^{aa}XX$ and $XX^{aa}XU^{a}XX$. These structures were determined by MALDI/CID MS/MS by Prof Toshihisa Kotake. Since CE4 removes Ac from mono-substituted Xyl, *m/z* = 1107 which has only two Ac, the only two possible oligosaccharides are $XU^{me}XX^{aa}XX$ and $XX^{aa}XU^{me}XX$. The same occurs for *m/z* = 1093, in which the only two possible structures are $XUXX^{aa}XX$ and $XX^{aa}XUXX$. Therefore, the lack of the Ac on glucuronidated Xyl would shift the signal from the species in red with three Ac to the ones with two Ac. If the lack of Ac is instead on di-acetylated Xyl, then CE4 would remove the second Ac, and GH11 would further digest the backbone, producing $XU^{[me],a}XX$ which are labelled in orange. In WT, the species in red with three Ac (*m/z* = 1135 and 1149) were predominant to the ones with two Ac (*m/z* = 1093 and 1107), meaning that the Xyl carrying a [Me]GlcA are mostly acetylated. *tbl32* produced higher proportions of the species with two Ac than species with three Ac ($Pent_6Ac_2U$ to $Pent_6Ac_3U$ and of $Pent_6Ac_2U^{me}$ to $Pent_6Ac_3U^{me}$) than the WT, denoting a partial loss of Ac on Xyl carrying a GlcA. In *tbl33* the signal of the species with two Ac (*m/z* = 1093 and 1107) were higher than *tbl32* and WT, and the signal of the species with three Ac (*m/z* = 1135 and 1149) were lower, indicating a stronger effect than in *tbl32*. The *tbl32/33* double mutant had a predominant production of the species in red with two Ac (*m/z* = 1093 and 1107), and the species with three acetates (*m/z* = 1135 and 1149) were close to the noise levels, which implies that this mutant has an almost complete loss of Ac in these glucuronidated Xyl sites. The effect exhibited by *tbl32*, *tbl33* and *tbl32/33* for the species in red are consistent with the

exhibited by the oligosaccharides in orange discussed above. In contrast, *esk1* showed almost no production of the species labelled with red, suggesting that in *esk1* these sites have no di-acetylated Xyl. This is consistent with the lack of *m/z* = 653 ($Pent_4Ac_2$ in brown), and the absence of $XX^{aa}XX$ on PACE discussed above.

In WT, the peaks of the oligosaccharides produced from acetylated Xyl that carry a methylated GlcA (*m/z* = 801 and 1149) were higher than the ones produced from the non-methylated GlcA (*m/z* = 787 and 1135), respectively. In *tbl32/33* these peaks were "replaced" by the oligosaccharides with one Ac less, because the Ac on glucuronidated Xyl is missing. Interestingly, the proportion of methylated GlcA (*m/z* = 759 and 1107) to non-methylated GlcA (*m/z* = 745 and 1093) is decreased. This evidence strongly suggests that the Ac transferred by TBL32 and by TBL33 impacts on the methylation of the GlcA on xylan.

The peaks with purple labels are produced by CE4/GH11 when $U^{[me],[a]}$ residues are spaced by three backbone Xyl from X^{aa}. Odd number of Xyl of distance between xylan substitutions have not been reported in the compatible domain or on primary-cell-wall xylan (Chapters 4 and 5), but they were detected in the incompatible domain of xylan (Chapter 4), therefore, the oligosaccharides labelled with purple are likely to originate from the incompatible domain. The oligosaccharides with seven Xyl produced by the WT have only three Acs; *m/z* = 1267 and 1281 ($Pent_7Ac_3U$ and $Pent_7Ac_3U^{me}$), with a higher proportion of the methylated species. *tbl32* produced the same two peaks, plus a species with two Ac; *m/z* = 1225 ($Pent_7Ac_2U$). Because of the activities of GH11 and CE4, in *m/z* = 1225 the Xyl carrying the GlcA is not acetylated. In *tbl33*, the methylated species with three Ac (*m/z* = 1281) is further reduced. *tbl32/33* produced almost none of the species with three Acs, and showed the highest production of species with only two Acs (*m/z* = 1225 and 1239). The different production of oligosaccharides with seven backbone Xyl by the mutants suggest that TBL32 and TBL33 are responsible for the 3-*O*-Ac on (methyl)glucuronidated Xyl present in the incompatible domain. Instead, *esk1* did not show any peaks with seven Pent. Which strongly suggests that ESK1 may be responsible for the di-acetylated Xyl in the incompatible domain.

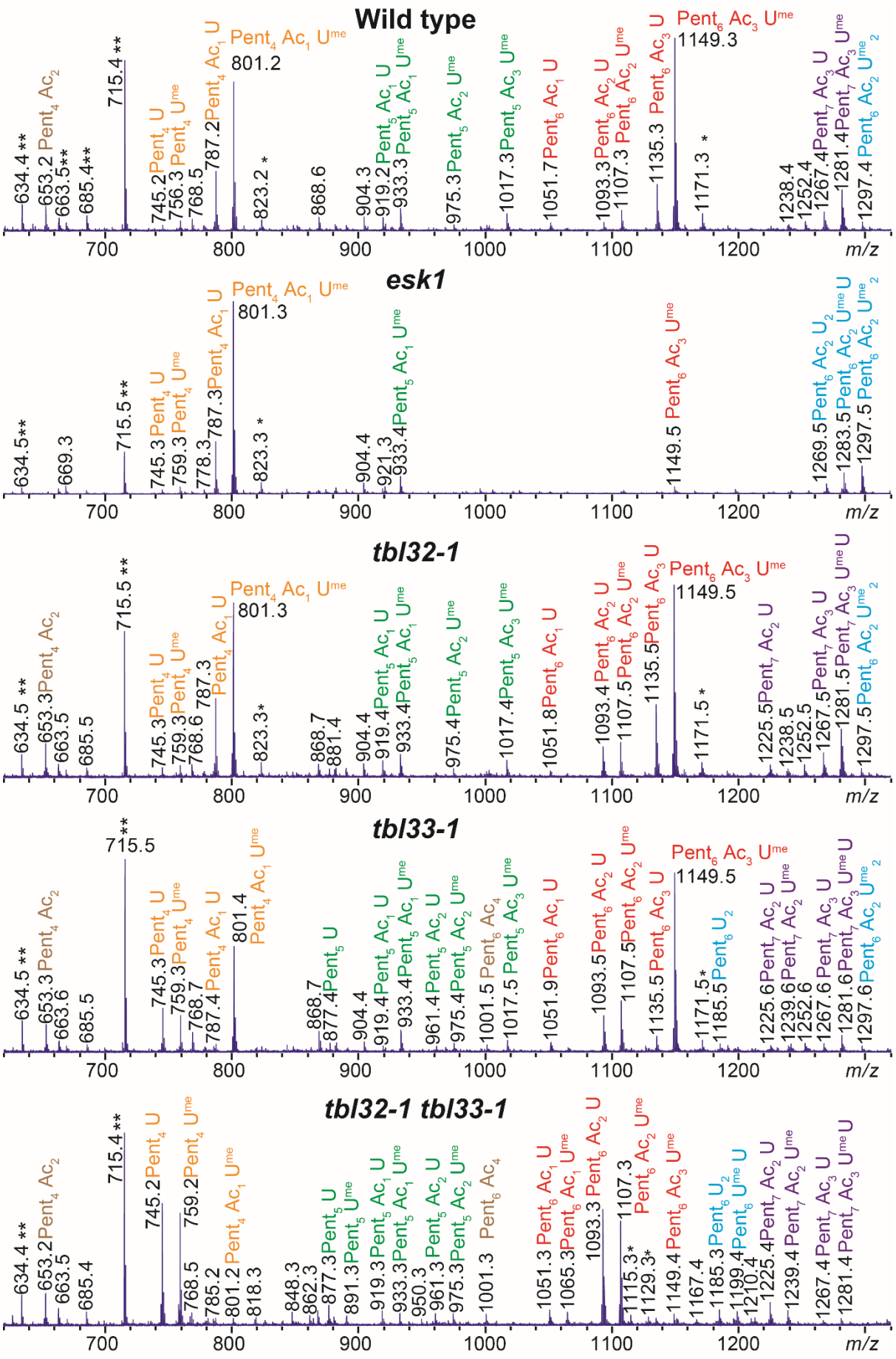
The origin of the oligosaccharides in green is uncertain, and their interpretation is challenging. These oligosaccharides have five Pent, and could originate from $U^{[me],[a]}$ residues of secondary cell wall xylan when next to X^{aa} in the incompatible domain

(Chapter 4). Alternatively, the activity of GH11 at high temperature (45 ºC) could perhaps produce a five Pent product when hydrolysing primary-cell-wall xylan with Ara*p* further modifications on GlcA (Chapter 5; Mortimer et al., 2015). The peaks *m/z* = 975 and 1017 ($Pent_5Ac_2U^{me}$ $Pent_5Ac_3U^{me}$) can only originate from the neighbouring of $U^{me,[a]}$ and X^{aa}, and were produced by WT, *tbl32* and *tbl33*. The lack of *m/z* = 1017 by *tbl32/33* suggests that TBL32 and TBL33 are responsible for the Ac on (methyl)glucuronidated Xyl in the incompatible domain. The peak *m/z* = 933 ($Pent_5Ac_1U^{me}$) was produced by WT and was not sensitive to the inactivation of ESK1 and TBL32/33. *M/z* = 933 is possibly originated from primary cell walls, and it would be interesting to determine its structure. Further interpretations of the production of species with five Pent would be ambiguous, and would be useful to analyse the structure using MALDI-ToF MS/MS.

The peaks labelled with blue arise from two $U^{[me],[a]}$ residues spaced by two backbone Xyl. In WT, *m/z* = 1297 ($Pent_6Ac_2U^{me}_2$,) was the only peak with two MeGlcA, which was close to the noise, denoting the low occurrence. *esk1* showed higher production of this species, and also produced *m/z* = 1269 and 1283 ($Pent_6Ac_2U_2$ and $Pent_6Ac_2U^{me}U$), which are the non-methylated GlcA versions of the former oligosaccharide. In *esk1*, the doubly-MeGlcA-substituted oligosaccharides carried two Ac, and because of CE4/GH11 activities, the Ac are likely to be on the MeGlcA-substituted Xyl. *tbl32* and *tbl33* have a similar *m/z* = 1297 peak height to WT, but in *tbl32/33* this species is absent. In both, *tbl33* and *tbl32/33,* the species $Pent_6U^{me}$ appears, and in the double mutant, also $Pent_6U^{me}U$ is present. The presence of these de-acetylated species supports TBL32 and TBL33 being responsible for the Ac on glucuronidated Xyl. The lack of methylation in the non-acetylated *m/z* = 1185 and 1199 supports that the Ac of which TBL32 and TBL33 are responsible for, impacts on the methylation of the GlcA.

The structure of the predominant CE4/GH11 products in the PACE analysis were identified by MS. The data presented here is strong evidence of that TBL32 and TBL33 are responsible for the acetylation of most of the glucuronidated Xyl. The impact on the GlcA methylation by the inactivation of TBL32 and TBL33 is contrary to the effect exhibited by the inactivation of ESK1. In addition, the MS spectrum of CE4/GH11 products on *esk1* xylan is strong evidence for the lack of di-acetylated xylan backbone residues in this mutant.

Figure 6.16: Mass-spectrometric analysis of GH11/CE4 digestion products of *tbl32/33* and *esk1* mutants' xylan.
Xylan was extracted from holocellulose prepared from mature bottom stems, and hydrolysed with GH11 xylanase and CE4 acetylesterase combined. The products were purified with a HyperSep column, mixed with DHB, and analysed on the mass-spectrometer.
The label colours classify the oligosaccharides by number of Pent and [Me]GlcA. Species in brown have 4 or 6 Pent and no [Me]GlcA. Species in red and blue have six Pent, and one (red) or two (blue) [Me]GlcA. Species in green and purple have one [Me]GlcA, and 5 or 7 Pent respectively. The species in orange have 4 Pent, one [Me]GlcA, and zero or one Ac. * doubly sodiated adducts. ** peaks produced by the matrix, DHB.


Wild type
esk1
tbl32-1
tbl33-1
tbl32-1 tbl33-1
m/z

The constant change in the ratio of species substituted with MeGlcA versus GlcA when the TBL32/33 Ac are present or absent suggests that the GlcA methylation process is driven by the Ac on the GlcA-substituted Xyl. As mass-spectrometry is not a fully quantitative technique, to calculate how much the methylation is affected by the Ac, a DASH approach was carried out as previously described (Li et al., 2013). The principle of this technique is similar to PACE but it has enough resolution to differentiate methyl groups. For this analysis, the single mutants *tbl32-1*, *tbl33-1*, *tbl33-2* and the double mutants *tbl32-1/33-1* and *tbl32-1/33-2* were used. In addition, with this approach it was possible to identify if the affected methylation on GlcA (and hence acetylation by TBL32/33) is xylan-domain-specific. To do this, extra mutants were generated; *tbl33-1 gux1* and *tbl33-1 gux2*, which only possess [Me]GlcA in either the incompatible or the compatible domains of xylan, respectively. These mutants were included in the analysis together with the controls *gux1* and *gux2*. The GlcA methylation levels vary depending on the portion of the stem analysed (Faria-Blanc, 2014), and for this reason, this analysis was done in triplicate by harvesting the lower 10 cm of mature inflorescence stems.

To carry out the DASH analysis, AIR was prepared, the xylan was extracted with alkali and completely digested with GH11. The products were labelled with the fluorophore APTS and resolved in a DNA sequencer as described (Li et al., 2013). The spectra obtained was analysed using DASHBOARD software (developed by Denis Rubtsov). The products XUXX and XU^{me}XX were the only products containing GlcA substitutions and were used to calculate the percentage of GlcA on xylan that are methylated. The values obtained were graphed and are shown in Figure 6.17.

The 70% of GlcA methylation in WT is in range to what has been reported before (Faria-Blanc, 2014; Grantham, 2016; Li et al., 2013; Peña et al., 2007). *tbl32* had very similar levels of methylation to WT, and both alleles of *tbl33* presented a slight reduction of about 15%. Both double mutants *tbl32/33* showed a drastic 50% reduction from WT levels. This is strong supporting evidence of that the lack of acetylation in glucuronidated Xyl has an effect on the methylation of the GlcA. The Ac by TBL32/33 is not necessary for GlcA methylation, but the data suggests that GXMs prefer the acetylated substrate during the xylan synthesis. Interestingly, the reduction observed in *tbl32/33* is more than the sum of both independent single mutants, which indicates that the substrates for both TBL32 and TBL33 are to some extent in common.

To analyse the domain-specificity of the effect of the Ac on the GlcA methylation, *gux1* and *gux2* were used as controls. The percentage of GlcA methylation in the *gux* mutants was around 80%, which is consistent with what was described by Grantham (2016) by using this method. The *tbl33 gux1* mutant had a reduction in GlcA methylation of about 36% respect to *gux1*, and *tbl33 gux2* had a 24% reduction respect to *gux2*. This suggests that the GlcA methylation in the compatible domain is less affected by the lack of TBL33 Ac than the incompatible domain. Because both *tbl33 gux1* and *tbl33 gux2* mutants showed a reduction of the percentage of GlcA methylation respect to *gux1* and *gux2*, it is implied that TBL33 is responsible for the acetylation of the compatible and the incompatible domains of xylan.

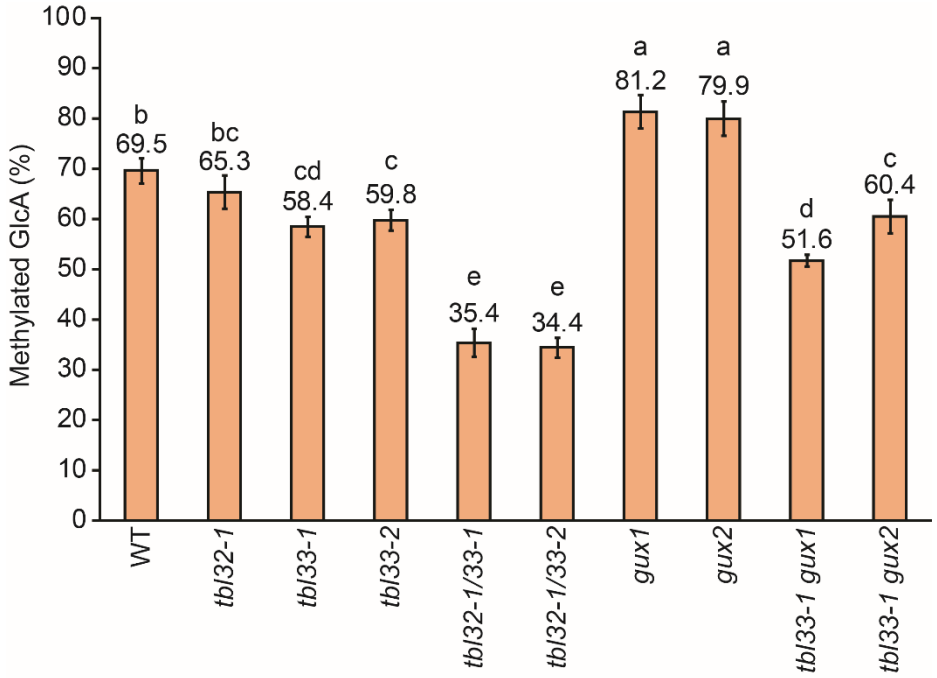


Figure 6.17: Methylation analysis of *tbl32*, *tbl33* and *tbl32/33* by GH11 digestion and DASH.
AIR was prepared from mature bottom stems, xylan was extracted with alkali and hydrolysed with GH11. The products were analysed by DASH using DASHBOARD software. The percentages were calculated from the abundance of the species XU^{me}XX and XUXX. The values are representative of three biological replicates. A Tukey-HSD test was implemented using the R software. Significant differences ($P < 0.05$) are indicated by different letters (a – e).

The PACE, MS and DASH approaches used in this section rely on the digestibility of xylan by GH11 xylanase. Therefore, possible non-digestible portions of xylan may exhibit an inconsistent behaviour to the observed on GH11 products, and the inactivation of TBL32 and/or TBL33 may not lead to a reduction of Ac on xylan and/or methylation of the GlcA in these indigestible regions. To analyse the acetylation and GlcA methylation on undigested xylan, NMR experiments were carried out. For this, xylan was extracted with DMSO from basal stems of WT, *tbl32-1*, *tbl33-1* and *tbl32-1/33-1*, and suspended in D_2O with 0.5 M urea. ^{13}C-HSQC spectra were obtained by Dr Teresa Almeida, and the peaks were assigned by Dr Teresa Almeida (Figure 6.18).

In panel A, the signal corresponding to $-CH_3$ in the Ac moiety linked to glucuronidated Xyl, was split into two; one arose from XG^{me}3 (see designations in Figure 6.18), and the other from XG3. These two peaks denote the 3-*O*-acetylation of (methyl)glucuronidated xylan Xyl. The peak volumes were calculated by Dr Teresa Almeida using AZARA software and are graphed in Figure 6.18 B. WT and *tbl32* produced both peaks at similar levels, *tbl33* produced both peaks in a minor magnitude, and in *tbl32/33* both peaks were absent. This shows that in *tbl33* xylan, the amount of acetylated U$^{[me],a}$ are reduced, but for the acetylated sites remaining, the proportion of methylated U^{me,a} to U^{a} is increased. In *tbl32/33* xylan no Ac was detected in glucuronidated sites. Curiously, *tbl33* showed a slight change in chemical shift of XG3, but not of XG^{me}3.

In Figure 6.18 panels C and D are shown the regions of the ^{13}C-HSQC spectra for H3C3 and H4C4 correlations of xylosyl residues. (Methyl)glucuronidated Xyl peaks were split into two, 3-*O*-acetylated (XG$^{[me]}$3) and non-acetylated (XG$^{[me]}$) in both panels. WT produced peaks corresponding to XG$^{[me]}$3 but not XG$^{[me]}$, indicating that all detectable (methyl)glucuronidated Xyl were 3-*O*-acetylated. The double mutant *tbl32/33* produced XG$^{[me]}$ peaks but not XG$^{[me]}$3, indicating that all detectable (methyl)glucuronidated Xyl in *tbl32/33* xylan were unacetylated. *tbl33* showed peaks of both XG$^{[me]}$3 and XG$^{[me]}$, denoting a partial loss of Ac in (methyl)glucuronidated Xyl. *tbl32* produced peaks of XG$^{[me]}$3, and not in XG$^{[me]}$. Both H3C3 and H4C4 peaks of XG$^{[me]}$3 in *tbl32* xylan were shifted for an unknown reason, and they were reduced compared to the WT. However, the acetylation levels of (methyl)glucuronidated Xyl in *tbl32* (panel B) were similar to the WT. The acetylation levels of glucuronidated Xyl exhibited by WT, the single mutants and the double mutant were consistent with the

analysis on PACE shown above. Therefore, this evidence supports the hypothesis that TBL32 and TBL33 are responsible for the acetylation of (methyl)glucuronidated Xyl, including the GH11-undigestible regions of xylan.

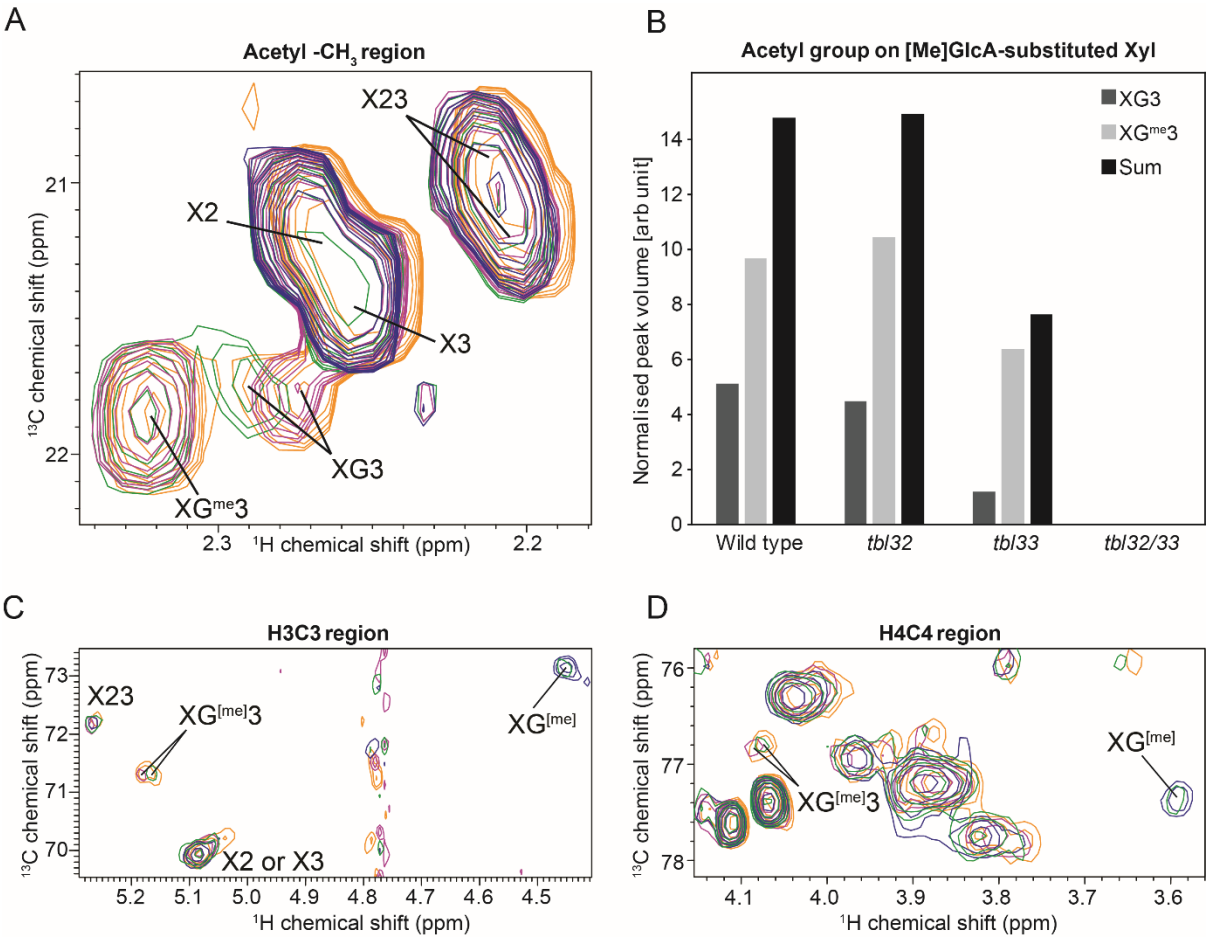


Figure 6.18: ^{13}C-HSQC NMR spectra.
AIR was prepared from basal stems, depectinated with 0.5% ammonium oxalate, delignified with 11% peracetic acid, and the xylan was extracted with DMSO at 80 °C. The DMSO was evaporated in vacuum, and the polysaccharides were resuspended in 3 M urea. The solubilised polysaccharides were purified by size exclusion with a PD10 column and vacuum-dried. The samples were resuspended in D_2O containing 0.5 M urea, and ^{13}C-HSQC spectra were acquired by Dr Teresa Almeida. X-axis: ^{1}H chemical shift, Y-axis: ^{13}C chemical shift, both in ppm. (A) $-CH_3$ peaks on *O*-acetyl groups on Xyl. (B) Peak volumes of XG^{me}3 and XG3 from panel A. Panels (C) and (D) show xylosyl H3C3 and H4C4 correlation regions respectively. Spectra for WT (orange), *tbl32-1* (pink), *tbl33-1* (green) and *tbl32-1/33-1* (blue) are shown. All peak assignments were carried out by Dr Teresa Almeida.
Designations modified from Teleman et al. (2000): X2, 2-*O*-acetylated Xyl; X3, 3-*O*-acetylated Xyl; X23, 2,3-di-*O*-acetylated Xyl; XG3, GlcA 2-*O*-linked and 3-*O*-acetylated Xyl; XG^{me}3, MeGlcA 2-*O*-linked and 3-*O*-acetylated Xyl; XG$^{[me]}$, [Me]GlcA 2-*O*-linked Xyl and XG$^{[me]}$3, [Me]GlcA 2-*O*-linked and 3-*O*-acetylated Xyl.

6.2.4.2 TBL28

The expression of TBL28 under the ESK1 promoter showed a restoration of *esk1* growth, but the [Me]GlcA pattern of xylan did not seem completely restored (section 6.2.2). The inactivation of ESK1 produces an alteration of the GlcA pattern on secondary-cell-wall xylan (Grantham et al., 2017), but none of the allelic mutants *tbl28-1* and *tbl28-2* mutants showed clear effects on primary- or secondary-cell-wall xylan GlcA pattern (section 6.2.3.3). Therefore, one hypothesis is that TBL28 has a similar function to ESK1, but TBL28 is less active. A second hypothesis is that TBL28 and ESK1 have a similar enzymatic activity, but they are differentially expressed. If they have a similar function, a simultaneous inactivation of TBL28 and ESK1 should lead to stronger phenotypes. However, the growth phenotype of *tbl28 esk1* double mutants was not stronger than of *esk1* (section 6.2.2.2).

To analyse the [Me]GlcA pattern in *tbl28 esk1* mutants, the plants were grown for six weeks and AIR was prepared from the lower portions of mature stems. The xylan was extracted with alkali and digested with GH30 glucuronoxylanase. The products were labelled with ANTS and analysed by PACE (Figure 6.19).

The WT xylan produced more $U^{[me]}X_8$ and $U^{[me]}X_{10}$ than $U^{[me]}X_7$ and $U^{[me]}X_9$ which is characteristic of the glucuronic acid by GUX1 (Chapter 4; Bromley et al., 2013). *esk1* produced higher amounts of $U^{[me]}X_6$ and $U^{[me]}X_7$ than the longer oligosaccharides, which is indicative of the altered [Me]GlcA pattern in this mutant (Grantham et al., 2017). The two *tbl28 esk1* double mutants produced more $U^{[me]}X_6$ and $U^{[me]}X_7$ than the oligosaccharides with longer chains, which is similar to the pattern exhibited by *esk1*. The production of X_2, X_3, X_4 and X_5 by all the samples is indicative of some breakage of some the backbone of some hydrolysis products. Considering this artefact, the similar abundance of all GH30 products suggests that *esk1* and the two *tbl28 esk1* mutants may have no clear difference in their GlcA pattern of xylan in secondary cell walls. Therefore, it is unlikely that TBL28 activity is related to the acetylation of xylan from secondary cell walls. However, it is possible that both TBL28 and ESK1 acetylate primary-cell-wall xylan, which would be interesting to analyse by investigating the GlcA pattern of xylan in primary walls in the double mutants. Also, because TBL28 is likely to be an acetyltransferase, it would also be useful to analyse the acetylation patterns of primary- and secondary-cell-wall xylan in the *tbl28 esk1* mutants.

In the analysis presented here it was not determined whether ESK1 and TBL28 are differentially expressed. However, this PACE analysis shows that the co-inactivation of TBL28 and ESK1 does not produce a stronger effect on the secondary-wall xylan [Me]GlcA pattern to the exhibited by *esk1*.

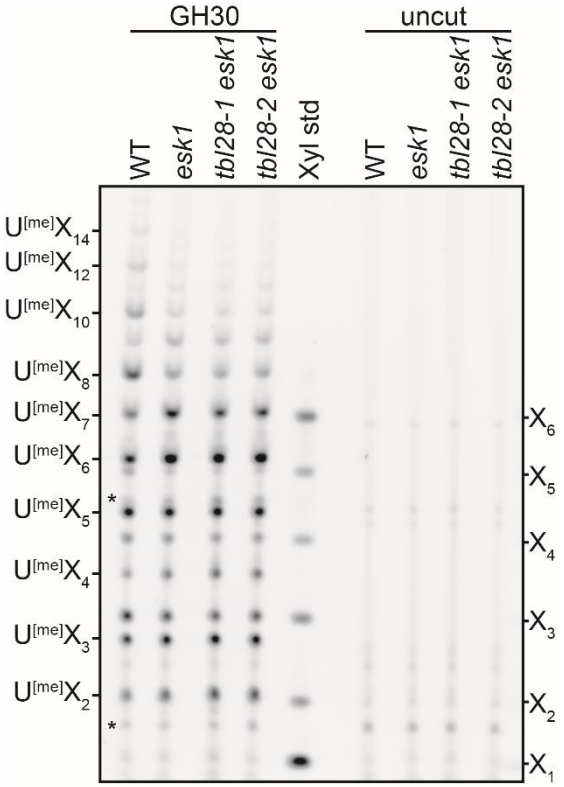


Figure 6.19: Analysis of GlcA substitutions spacing on xylan of *tbl28 esk1* mutants by PACE.
AIR was prepared from mature bottom stems of 6-week-old plants. Xylan was extracted with alkali and digested with BoGH30. The oligosaccharides were labelled with ANTS and resolved on a 10% PACE gel. This analysis was carried out for one biological replicate for both allelic *tbl28 esk1* lines.

6.2.4.3 AXY9

It was proposed that the function of AXY9 is to provide the acetylated donor for the TBLs for the acetylation of multiple cell wall polysaccharides (Schultink et al., 2015). Consistent with this hypothesis, it was shown in section 6.2.3.3 that the *axy9* mutant has an altered [Me]GlcA pattern of xylan, which is a similar effect to that produced by the inactivation of the xylan acetyltransferase ESK1 (Grantham et al., 2017). The inactivation of AXY9, however, may affect xylan acetylation by other acetyltransferases. To identify whether the inactivation of AXY9 affects only ESK1, or

also other xylan acetyltransferase activities, the xylan acetylation of *axy9* and *esk1* were analysed and compared.

The PACE analysis of GH11 products (section 6.2.3.4) showed that *esk1* xylan is more accessible to the xylanase than WT, because this mutant lacks Ac in the alternating acetylation pattern. Also, no di-acetylated Xyl were detected in *esk1* xylan by PACE and MS using GH11/CE4 (sections 6.2.3.5 and 6.2.4.1). Because ESK1 activity may be affected in *axy9*, gaps in the alternating acetylation pattern and loss of di-acetylated Xyl are possible. To investigate which Ac are missing in *axy9*, its xylan was analysed with the two methodologies used previously for the analysis of the *tbl* mutants this chapter (sections 6.2.3.4 and 6.2.3.5). The results of both analyses on PACE is described next.

First, to detect gaps on the alternating acetylation pattern on xylan, the accessibility of *axy9* xylan to GH11 was tested. The hydrolysis products were either treated or not with alkali, labelled with ANTS and analysed on PACE (Figure 6.20 panel A). The production of X_1 and X_2 by WT xylan shows that GH11 had some access to the xylan. Higher accessibility to *esk1* xylan by GH11 is implied by the higher production of X_1, X_2, and $XU^{[me],a}XX$, as seen in section 6.2.3.4. GH11 hydrolysis of *axy9* xylan produced a higher amount of X_1 and X_2, and produced the non-acetylated $XU^{[me]}XX$ oligosaccharide instead of $XU^{[me],a}XX$ (blue arrow). After the alkali treatment, $XU^{[me],a}XX$ produced by WT and *esk1* was de-acetylated, and co-migrated with $XU^{[me]}XX$ produced by *axy9*, which confirms the identity of both oligosaccharides. In section 6.2.4.1 was shown that TBL32 and TBL33 are responsible for all 3-*O*-Ac on Xyl that carry a [Me]GlcA. Since GH11 hydrolysis of *axy9* xylan produced $XU^{[me]}XX$ instead of $XU^{[me],a}XX$, it is implied that the inactivation of AXY9 has an impact on the activity of TBL32 and TBL33. Some GH11→alkali products on WT xylan (blue asterisks) were not produced by *esk1* and *axy9*, and one product was produced by *esk1* and *axy9* and not by WT (red asterisk). A similar behaviour of these products between *esk1* and *axy9* suggests that these oligosaccharides are generated when ESK1 activity is affected. However, one GH11→alkali product (green asterisk) was produced by WT and *esk1* but not by *axy9*, which suggests that this oligosaccharide is produced when an acetyltransferase other than ESK1 is affected.

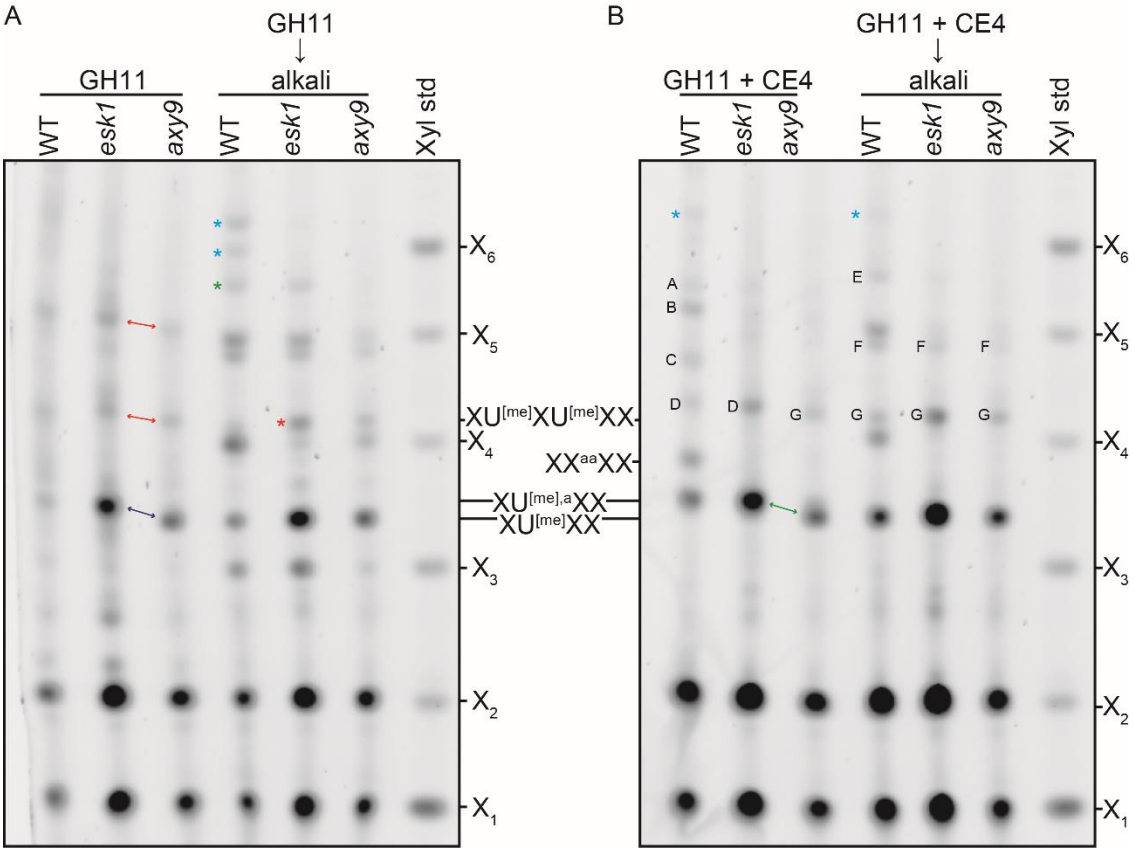


Figure 6.20: Analysis of xylan acetylation of *esk1* and *axy9* by PACE.
6-week-old plants were grown. *esk1* plants were grown on soil and *axy9* on magenta pots in sterile conditions. Basal mature stems were collected from *esk1* and whole stems from *axy9*, and holocellulose was prepared. The samples were digested with GH11 (panel A) or with GH11 and CE4 (panel B). The enzymes were inactivated with heat and removed using Nanosep. The hydrolysis products were then treated with alkali, labelled with ANTS and resolved on 10% PACE gels.
Arrows: Red arrows show difference in migration of unidentified acetylated species. Blue and green arrows show the difference in migration between $XU^{[me],a}XX$ and $XU^{[me]}XX$.
Asterisks: Blue asterisks show bands produced by the digestion of WT xylan but not by *esk1* and *axy9*. Red asterisk shows a band produced by *esk1* and *axy9* but not by WT. Green asterisk shows a band produced by WT and *esk1* but not by *axy9*.
Letters: Bands A-G are coherent with the makings in a similar PACE analysis shown in section 6.2.3.5, and a summary of the putative identities is presented in Table 6.2 (page 212). Bands A-D are unidentified acetylated species. Bands E-G are produced by acetylated species when treated with alkali.

Second, to identify Ac deficiencies associated with di-substituted Xyl, holocellulose from *axy9*, *esk1* and WT was incubated with CE4 and GH11 simultaneously. The hydrolysis products were either treated or not with alkali, then they were labelled with ANTS and analysed on PACE (Figure 6.20 panel B). WT and *esk1* produced $XU^{[me],a}XX$, while *axy9* produced $XU^{[me]}XX$ (green arrow). This suggests that the inactivation of AXY9 impacts on the activity of TBL32 and TBL33, which supports the evidence from GH11 alone (panel A, blue arrow). The oligosaccharide $XX^{aa}XX$ is produced by GH11/CE4 from xylan regions with di-acetylated Xyl that are isolated from other R substitutions ($U^{[me],[a]}$ or X^{aa}). In section 6.2.4.1 was shown that $XX^{aa}XX$ was produced by WT and was undetectable on *esk1* products by MS. Here, $XX^{aa}XX$ was only produced by WT, and was undetectable by PACE on *esk1* and *axy9* products. This is consistent with the view that in *axy9*, the alteration of ESK1 activity affects the production of di-acetylated Xyl. It could be that a second acetyltransferase involved in di-acetylation requires the activity of AXY9. The absence of species A, B and C in both *axy9* and *esk1*, also suggests that ESK1 activity is required for the production of these oligosaccharides.

The results here shown are supporting evidence for AXY9 being responsible for several Ac on xylan. This included the Ac mediated by ESK1 and TBL32/33. The identification of the remaining Ac in *axy9* GH11 and GH11/CE4 products by MS would be useful for determining whether AXY9 has a preference for a specific xylan acetyltransferase activity.

6.3 Discussion

In Chapters 4 and 5 was shown that different xylan regions have different patterns of acetylation. In the xylan compatible domain, Xyl substituted with single Ac were alternated with unsubstituted Xyl, and in the incompatible domain several acetylated Xyl were found successively arranged. Di-acetylated Xyl, and Xyl di-substituted with one Ac and an additional modification ([Me]GlcA, PUX or Ara*f*) were also described. The multiple positions and patterns of Ac on xylan therefore suggests that several xylan acetyltransferases are possible.

The main aim of this chapter was to identify the acetyltransferases that are responsible for the multiple Ac positions on xylan. The phylogenetic analysis carried out in this chapter suggested that within the ESK1 homologues in eudicots, between five or six putative activities are conserved. *Arabidopsis* possesses nine sequences in this clade, with either one or two genes for each possible activity. Since the activities of this clade were shown to be responsible for more than one Ac position on xylan, it is likely that all the activities in the ESK1 branch are associated with xylan. Each of the possible functions found here is discussed in the following sections.

6.3.1 ESK1 acetylation determines other downstream xylan acetylation and glucuronidation processes

ESK1 was shown to be responsible for most mono-Ac substitutions on xylan. The higher accessibility to xylan by GH11 xylanase exhibited on the PACE analysis of *esk1* acetyl-xylan (section 6.2.3.4) indicates that the xylan in this mutant lacks some of the modifications on the backbone. This is supported by the digestion with GH11 and CE4 combined (Figure 6.14 and Figure 6.20B), where the production of $XU^{[me],a}XX$ by WT is closer to the levels of *esk1* mutant. WT xylan accessibility by GH11 is enhanced by the removal of mono-Ac-substitutions by CE4, but since WT xylan has X^{aa} and *esk1* does not, WT xylan is less accessible by CE4/GH11 in regions where X^{aa} occur. This evidence is consistent with the NMR data reported by Grantham (2016), in which the ^{13}C-HSQC spectra of *esk1* shows undetectable X2 (2-*O*-Ac) cross-peaks and small amounts of X3 (3-*O*-Ac) compared to WT. Other reports also showed by solution-state 1-D ^{1}H-NMR experiments that *esk1* xylan has a reduction in both 2-*O*-Ac and 3-*O*-Ac (Xiong et al., 2013; Yuan et al., 2013). Nevertheless, it is unclear whether 2-*O*-Ac and

3-*O*-Ac are similarly reduced in xylan in the cell wall. Since the analysis of extracted xylan by NMR allows enough time for Ac to migrate before the data is acquired, the loss of 3-*O*-Ac in the spectrum might not reflect loss of 3-*O*-Ac of the xylan in the cell wall. The ESK1 *in vitro* activity reported by Urbanowicz et al. (2014) is also consistent with the results shown in this chapter. In the report xylohexaose was acetylated by ESK1 at O-2, and only after 30 min of incubation 3-*O*-acetylated products were detected by real-time 2-D NMR, which is strong evidence of ESK1 to be regiospecific for 2-*O*-Xyl acetylation.

The lack of O-2-specific Ac is perhaps reflected in the activity of GH11 on *esk1* acetyl-xylan. It was proposed in section 6.2.3.4 that X_3Ac_1 and X_4Ac_1 are produced by GH11 when it hydrolyses regions of xylan with isolated mono-acetylated Xyl (Figure 6.13 and reproduced in Figure 6.20A). Then, when GH11 products were de-acetylated with alkali, X_3 and X_4 were produced respectively. Alternatively, X_4 could be produced by the de-acetylation of $XX^{aa}XX$. However, because no di-acetylated oligosaccharides were detected in *esk1* xylan (Figure 6.14 and Figure 6.16), all X_4 produced by GH11→alkali originated from X_4Ac_1. It is unclear why in *esk1* the production of X_3 was increased, and X_4 was decreased. GH11 xylanases do not tolerate substitutions at the -1 or +1 subsites, but have been shown to tolerate Ara*f* at the -2 on the 3-OH but not the 2-OH (Ridlova et al., 2008). It is possible that X_3 and X_4 are generated by GH11→alkali when the enzyme hydrolyses the backbone around a 3-*O*- or 2-*O*-mono-acetylated Xyl respectively. Then, because ESK1 is 2-*O*-Ac regiospecific (Urbanowicz et al., 2014), the production of X_4 is decreased in this mutant possibly because the missing Ac are likely to be O-2-linked. Then, X_3 is increased possibly because of the higher accessibility by GH11 to *esk1* xylan (discussed above). Spontaneous migration of Ac between O-2 and O-3 of backbone Xyl has been previously reported to occur on xylan in solution (Kabel et al., 2003; Mastihubová and Biely, 2004; Urbanowicz et al., 2014), however, here the GH11 hydrolyses were carried out on holocellulose, in which xylan is naturally adsorbed and not in solution. This difference could explain the difference in GH11 activity observed between WT and *esk1* xylan. In addition, the GH11→alkali treatment of *axy9* holocellulose also is consistent with the hypothesis of the Ac-position-dependent GH11 activity. AXY9 has been proposed to be an intermediary between RWA and the TBLs (Schultink et al., 2015), and its inactivation depletes ESK1 and other TBLs of Ac donor. Therefore, it was expected to find a

reduction of ESK1-mediated Ac, but also of other-TBL-mediated Ac. The GH11→alkali treatment products X_3 and X_4 by *axy9* holocellulose were both decreased respect to the WT (Figure 6.20A), but the higher proportion of X_4 was maintained. If GH11 activity is dependent of the Ac position, this would suggest that 2-*O*-Ac and 3-*O*-Ac are both reduced in *axy9* xylan. It would be useful to determine whether GH11 activity on xylan is different when digesting holocellulose in comparison to DMSO-extracted acetyl-xylan.

ESK1 also was found to be absolutely necessary for di-acetylated Xyl synthesis. The CE4/GH11 on *esk1* xylan produced a complete absence of $XX^{aa}XX$ on PACE (section 6.2.3.5), which is indicative of the lack of di-acetylated Xyl in this mutant. Supporting evidence was shown by MS, which confirmed the lack of $XX^{aa}XX$ in CE4/GH11 products, and also showed no di-acetylated Xyl near [Me]GlcA xylan substitutions (section 6.2.4.1). This is supported by the NMR data reported by Grantham (2016), in which the ^{13}C-HSQC spectra of *esk1* xylan in which the cross-peaks corresponding to di-acetylated Xyl (X23) were absent. Several publications from one group during the course of the work for this thesis have reported by 1-D ^{1}H-NMR experiments of *esk1* xylan, and they interpret that ESK1 has no effect on X^{aa} residues (Yuan et al., 2016a, 2016c, 2016b, 2013). The frequency of X^{aa} on xylan is however inconsistent between these publications for WT and for *esk1*. These inconsistencies are likely to arise when using 1-D ^{1}H-NMR methods for quantification in complex samples. As there is no separation in the carbon-dimension, it is very likely that they considered unresolved acetylation-unrelated peaks in the measurements. Xiong et al. (2013) described *esk1* xylan acetylation by 2-D NMR experiments, however, the spectra of the samples reported were too complex to resolve di-acetylated Xyl from mono-acetylated Xyl. Three hypotheses therefore arise; First, ESK1 may have two acetyltransferase activities, one 2-*O*-acetyltransferase and another 3-*O*-acetyltransferase. Because transferase enzymes commonly have a single regiospecificity, this hypothesis is highly unlikely. However, this first hypothesis is perhaps strictly possible because there are reports of some transferases that can modify sugar residues with more than one different linkage. One example is CSLF, which elongates the backbone of mixed-linked glucans and has been shown to transfer 1,3 and 1,4-glucosyl residues to the backbone (Burton et al., 2006). A second hypothesis is that ESK1 2-*O*-acetylates a backbone Xyl, then this Ac spontaneously migrates to O-3 while xylan is in solution in

the Golgi, and this would allow ESK1 to re-acetylate the Xyl on O-2. A third hypothesis is that the Xyl that are 2-*O*-acetylated by ESK1 are substrate for a downstream 3-*O*-acetyltransferase which is responsible for the second Ac. This third hypothesis is most likely to be true because it is the simplest explanation. However, the activity of this hypothetical ESK1-activity-dependent xylan-acetyltransferase was not identified in the screen for phenotypes of TBL mutants.

Xylan acetylation by ESK1 determines the patterning of xylan with GlcA. The PACE analysis of *esk1* using GH30 (section 6.2.3.3) showed an altered GlcA pattern of xylan, which is consistent with the reported by Grantham et al. (2017). The [Me]GlcA spacing in *esk1* xylan showed no preference for even number of backbone Xyl, which is indicative of the altered [Me]GlcA pattern in this mutant. Supporting evidence for the altered pattern of [Me]GlcA was shown by MS on CE4/GH11 products (section 6.2.4.1), in which *esk1* showed a higher production of xylooligosaccharides carrying two [Me]GlcA modifications. This implies that when ESK1 is inactivated, GUX1/2 are dysregulated, and the [Me]GlcA on xylan are more frequently at short distance from each other. Naturally-occurring consecutive Xyl carrying MeGlcA have been recently reported in spruce xylan (Martínez-Abad et al., 2017), however, in *esk1* xylan such short distance between two [Me]GlcA was not detected, since the minimum distance observed between [Me]GlcA-modified Xyl was of two backbone residues.

The data shown here implies that the function of ESK1 is necessary not only for the main alternating acetylation pattern of xylan, but also for the acetylation of other sites on xylan and for the patterning of the GlcA. How ESK1 activity during the xylan synthesis determines these other downstream processes is not yet understood. However, since ESK1 is likely to be only an O-2-specific xylan acetyltransferase, one hypothesis is that ESK1 catalyses the first 2-*O*-acetylation on xylan, and that these Ac are transferred to every other Xyl at that stage. Since GlcA by the GUX enzymes and Ac by ESK1 are both O-2-linked to Xyl, two possibilities arise; ESK1 may leave residues undecorated. Alternatively, the acetylation pattern by ESK1 is adapted after by acetyltransferases, transacetylases and/or acetylesterases or by migration to O-3 in order for GUX1 and GUX2 to recognise their glucuronidation sites. Both possibilities are consistent with Grantham et al (2017), who proposed that the GUX enzymes may recognise “gaps” in the ESK1 acetylation pattern of xylan. For testing both possibilities,

it would be useful to identify whether the transfer of Ac by ESK1 is carried out in a patterned manner. This could be tested by sequencing oligosaccharides decorated *in vitro* with two Ac by ESK1 using MALDI-ToF MS/MS. A second possible hypothesis is that GUX1 and GUX2 regiospecificities recognise more complex acetylation arrangements than just gaps in the alternating acetylation pattern of xylan, such as di-acetylated Xyl or consecutive mono-acetylated Xyl. If this hypothesis is true it would be necessary to free up the 2-OH for additional xylan modifications to happen after xylan is acetylated by ESK1 and before glucuronidation by GUX1/2 occurs. Evidence for complex acetylation patterns on regions of xylan that are decorated by GUX1 and GUX2 is shown in the MS analysis of xylan from *gux1* and *gux2* in Chapter 4. Also, the analysis of spacing between [Me]GlcA on xylan from *gux1* and *gux2* presented by Bromley et al (2013) suggests that the activity of GUX1 and GUX2 are restricted to decorate independent domains of xylan. In agreement with both hypotheses presented here, they proposed that GlcA decorations by GUX1 and GUX2 may have a different acetylation context. Since in *esk1* xylan the main alternating pattern of acetylation is missing, the dysregulation of GUX1 and possibly GUX2 may be caused because new acetylation patterns can be recognised by GUX1 and GUX2.

6.3.2 TBL28 can partially complement the lack of ESK1

The phylogenetic analysis of ESK1 homologues in higher plants revealed that TBL28 and TBL30 are likely to have a similar activity to ESK1 (section 6.2.1.2). This is because they both grouped with a single sequence of the early-diverging angiosperm *Amborella*, which suggests a conservation of a single enzymatic activity among angiosperms. Since *TBL28* and *TBL30* are adjacent genes in *Arabidopsis* genome (At2G40150 and At2G40160) they likely originated from a recent gene duplication rather than due to a convergent evolutionary process. However, because the topology of TBL30 is isolated from other eudicots in the ESK1 sub-clade, TBL30 is likely to have a different activity to TBL28 and ESK1.

Since the expression patterns of *TBL28* and *TBL30* are similar to GUX3, possibly TBL28 and TBL30 may be involved in the synthesis of xylan in primary cell walls. The inactivation of ESK1 showed an effect on the substitutions on xylan (section 6.2.3; Grantham et al., 2017), therefore it was possible that the inactivation of TBL28 or

TBL30 could cause a similar effect on primary-cell-wall xylan. As discussed above, the lack of ESK1 causes a deficit of secondary-cell-wall xylan acetylation and disrupts the patterns of Ac and [Me]GlcA on the backbone. This causes a mildly dwarfed growth effect on *esk1* mutant because the xylan is less functional. However, the growth of *tbl28* and *tbl30* mutants was unaffected, and their GlcA patterns on primary- and secondary-cell-wall xylan were similar to the WT. This suggested that the functions of TBL28 and TBL30 are different to ESK1. Supporting evidence is shown by the lack of additional effect on the growth of *esk1* of the inactivation of TBL28 and TBL30 in the double mutants *tbl28 esk1* and *tbl30 esk1*. In addition, *tbl28 esk1* showed no further alteration of the [Me]GlcA pattern on PACE. On the other hand, it was also possible that TBL28 and TBL30 may have redundant functions in primary cell wall, then it would be useful to analyse the effects of the simultaneous inactivation of TBL28 and TBL30. Because ESK1 may also be involved in this process it would be also useful to investigate the primary-wall xylan acetylation of *esk1 tbl28/30* triple mutant.

It is possible that TBL28 or TBL30 may have a similar enzymatic activity to ESK1, therefore, to test whether TBL28 and TBL30 can substitute the activity of ESK1 they were expressed in *esk1* mutant under the promoter of *ESK1* (section 6.2.2). The overexpression of TBL28 in secondary walls restored the dwarfed growth caused by the inactivation of ESK1, but *pESK1:TBL30*(*esk1*) lines showed no clear effect on the growth. The PACE analysis using GH30 of the TBL28 overexpressor lines showed a marginal recovery of the [Me]GlcA pattern of xylan, which is supporting evidence of that TBL28 has a similar activity to ESK1. During the course of the work for this thesis, Zhong et al. (2017) reported a recovery of *esk1* growth when overexpressing TBL28 which is consistent with the result of the promoter swap experiment shown here. For the overexpression they used the promoter of *CESA7* which is approximately four times stronger than the promoter of *ESK1* used here (TRAVA, RNAseq data).

Zhong et al. (2017) recently showed that TBL28 and TBL30 heterologously expressed in human HEK293F cells and purified are catalytically active on xylohexaose, by detecting a mass difference of the products corresponding to one or two Ac. In agreement with the hypothesis stated above, they proposed that ESK1 and TBL28 may have a similar activity, because the chemical shift of the 2-*O* and 3-*O*-acetylated products in a 1-D ^{1}H-NMR experiment have a similar signal ratio. They also propose

that TBL30 has a different activity because the ratios are different and lacks an additional peak, which they believe corresponds to di-acetylated Xyl (H3-X23). However, another publication used 2-D NMR experiments to monitor the ESK1 acetyltransferase activity to xylooligosaccharides, and the various acetylated backbone Xyl had different chemical shifts (Urbanowicz et al., 2014). Since the chemical shift assignments by Zhong et al. (2017) are based only on previously reported 1-D NMR experiments, and because acetylated Xyl in different context may produce different shifts, their NMR analysis requires a second (^{13}C) dimension to verify the identity of the peaks in their spectra. It would be interesting to investigate whether the Ac by TBL28 and TBL30 on xylan products are at different positions.

6.3.3 Acetylation of glucuronidated Xyl of xylan is completely mediated by TBL32 and TBL33, occurs in regions decorated by GUX1 and GUX2, and impacts on the downstream process of GlcA methylation

Since the 3-*O*-acetylation of 2-*O*-[Me]GlcA substituted Xyl is unaffected by the inactivation of ESK1 (MS analysis in section 6.2.4.1; Grantham, 2016; Yuan et al., 2013), it is implied that another acetyltransferase activity is required for the acetylation of these sites. The PACE analysis using CE4/GH11 (section 6.2.3.5) was used to screen for acetylation deficiencies on xylan of the *tbl* mutants. TBL33 was found to be partially responsible for the acetylation of glucuronidated Xyl on xylan, because the xylan hydrolysis of both allelic mutants *tbl33-1* and *tbl33* produced the non-acetylated species $XU^{[me]}XX$ in addition to $XU^{[me],a}XX$. No other *tbl* single mutant analysed had this phenotype, but the phylogenetic analysis on higher plants suggested that TBL32 and TBL33 may have a similar activity, since both grouped with a single sequence from the early-diverging angiosperm *Amborella*. Then, to test if TBL32 contributes to the acetylation of glucuronidated Xyl, the double mutant *tbl32/33* was produced and its xylan was analysed and compared to the single mutants. The analysis of *tbl32/33* and both single mutants by PACE and MS using CE4/GH11 (section 6.2.4.1) revealed that TBL32 and TBL33 are both necessary for the specific acetylation of glucuronidated Xyl on secondary-cell-wall xylan. The levels of acetylation in glucuronidated Xyl were barely undetectable in the double mutant on either PACE and MS, and because there are no other *Arabidopsis* members in the TBL32/33 subclade, it is very likely that only TBL32 and TBL33 are responsible for the acetylation of all of

glucuronidated Xyl. In contrast, the PACE and MS analyses showed no effect on the production of $XX^{aa}XX$ by either *tbl32*, *tbl33* or *tbl32/33*, which implies that neither TBL32 nor TBL33 is responsible for the production of di-acetylated Xyl. Supporting evidence for the Ac position that is sensitive to mutations in TBL32 and TBL33, is shown by the GH11/CE4 products $Pent_6Ac_2U^{[me]}$ and $Pent_6Ac_3U^{[me]}$ which indicate the absence or presence (respectively) of Ac on glucuronidated Xyl. In *tbl32* and *tbl33* the production of these oligosaccharides was shifted towards $Pent_6Ac_2U^{[me]}$, while *tbl32/33* generated almost only $Pent_6Ac_2U^{[me]}$, denoting a partial loss of Ac in the single mutants, and a more severe deficit on the double mutant xylan, but Ac only on glucuronidated Xyl and not on di-acetylated or mono-substituted Xyl. Because PACE and MS analyses are based on the digestible fractions of xylan, supporting evidence for indigestible regions is shown in the ^{13}C-HSQC spectra of *tbl32/33* xylan, which shows no detectable Ac on Xyl carrying a [Me]GlcA. The PACE, MS and NMR data presented in this chapter shows strong evidence of that TBL32 and TBL33 are responsible uniquely for the 3-*O*-acetylation of secondary-wall xylan glucuronidated Xyl, and therefore, both proteins are likely to be glucuronoxylan-3-*O*-acetyltransferases.

During the course of the work for this thesis, Yuan et al. (2016a) reported an MS analysis of *tbl32/33* xylan using a GH11 xylanase. Since they did not use xylan acetylesterase CE4 for the hydrolysis, the accessibility of GH11 to the xylan was restricted and the interpretation of their data is narrowed to a smaller digestible fraction of the xylan compared to the analyses shown in this chapter. However, their MS spectrum of *tbl32/33* xylan digested with GH11 consistently shows a higher production of species with less acetate than the WT. In agreement with the hypothesis stated here, they concluded that the activity of TBL32/33 is different to ESK1 because the oligosaccharides produced by GH11 on *tbl32/33* and *esk1* xylan are different. However, they did not show the independent contribution of TBL32 and TBL33 for the acetylation of these sites. To show the position of the lacking Ac on *tbl32/33* xylan Yuan et al. (2016a) presented a 1-D ^{1}H-NMR spectrum, and integrated the area of a group of about six overlapped peaks in the spectrum according to the proton resonance of the H3 of glucuronidated Xyl (between 5.05 and 5.2 ppm). Based on their NMR results they mention that the xylan of the *tbl32/33* mutant possesses barely undetectable levels of Ac in glucuronidated Xyl. Nevertheless, since the analysis by

NMR of a polymer with heterogeneous substitutions such as xylan is highly complex, it is not possible to determine the identities of each of the peaks that they used for the integration only by using the ^{1}H dimension.

In another recent publication by the same research group (Zhong et al., 2017) they tested purified TBL32 and TBL33 for acetyltransferase activity on X_6 and XUXX. Their 1-D ^{1}H-NMR analysis of the enzymatic reaction products suggested that TBL32 and TBL33 are only active on the glucuronidated substrate by producing a single peak at 5.15 ppm corresponding to the H3 of 3-*O*-acetylated 2-*O*-GlcA-substituted Xyl. The enzymatic activities proposed by Zhong et al. (2017) for TBL32 and TBL33 support the genetic analyses presented here, where the independent inactivation of TBL32 and TBL33 exhibit deficits of Ac on glucuronidated Xyl.

The lack of 3-*O*-Ac by TBL32/33 on 2-*O*-[Me]GlcA-substituted Xyl showed an effect on the level of methylation of the GlcA. In WT xylan the MS analysis using GH11 and CE4 (section 6.2.4.1) showed a higher production of oligosaccharides with a methylated GlcA than the respective analogue oligosaccharides with a non-methylated GlcA. This is consistent with several previous glucuronoxylan MS analyses reported (Bromley et al., 2013; Chong et al., 2014; Grantham, 2016), and it reveals that in *Arabidopsis* not all GlcA modifications on xylan are methylated. In the MS analysis shown here, the mutant *tbl32/33* exhibited a reduced proportion of oligosaccharides with a MeGlcA to the analogue ones with a non-methylated GlcA. Also, in the *tbl33* mutant, where only part of the 2-*O*-[Me]GlcA-substituted Xyl are 3-*O*-acetylated, the ratio of MeGlcA:GlcA was maintained for the fully acetylated oligosaccharides, and it was reduced in the Ac-deficient oligosaccharides. This evidence shows there is an impact on the downstream process of GlcA methylation by the GXM methyltransferases driven by the Ac by TBL32/33.

However, in *tbl32/33*, where glucuronidated Xyl are completely depleted of Ac, the GlcA modifications on xylan can still be methylated, which indicates that 3-*O*-Ac on GlcA-Xyl may facilitate the GlcA methylation by producing a better substrate for the GXM.

Supporting evidence for the reduction in GlcA methylation is shown by the quantitative method DASH using GH11 (section 6.2.4.1; Li et al., 2013). The percentage of methylated GlcA in WT was of 70% which is in range to what has been previously

reported (Bromley et al., 2013; Chong et al., 2014; Faria-Blanc, 2014; Grantham, 2016; Li et al., 2013; Peña et al., 2007). The *tbl32* mutant exhibited a non-significant 6% reduction of GlcA methylation from the WT, both allelic mutants *tbl33-1* and *tbl33-2*, which have less Ac than *tbl32-1*, showed a significant 14% and 16% reduction of GlcA methylation from the WT respectively, and drastically the two double mutants *tbl32-1/33-1* and *tbl32-1/33-1*, which are depleted of Ac on 2-*O*-[Me]GlcA-substituted Xyl, showed a significant 49% and 50% reduction of GlcA methylation from WT. Therefore, this evidence shows that the level of GlcA methylation on xylan directly correlates to the DA on the 2-*O*-[Me]GlcA-substituted Xyl. Since in WT almost 100% of Xyl carrying a [Me]GlcA are acetylated (Chong et al., 2014), and *tbl32/33* mutants specifically lack these Ac, it implies that TBL32- and TBL33-mediated Ac impact on the methylation of half of the GlcA on xylan. In addition, since both *tbl33-1/gux1* and *tbl33-1/gux2* exhibited a reduction in GlcA methylation from the respective controls (*gux1* and *gux2* respectively), it is implied that the Ac by TBL32 and TBL33 impact on the degree of GlcA methylation in the compatible and incompatible domains of xylan.

Other xylan acetylation mutants such as *esk1* and *rwa* have also been shown to have altered methylation levels on xylan GlcA substitutions (Grantham, 2016; Grantham et al., 2017; Manabe et al., 2013). RWA have been predicted to be transporters for the import of Ac-CoA into the Golgi, because they are type III membrane proteins and because their inactivation affects the acetylation of several polymers of the cell wall. In the *rwa1/2/3* and *rwa1/3/4* triple mutants, where only one functional RWA protein remains, xylan acetylation is reduced by 20-30% and the GlcA methylation is decreased by about 13% respect to the WT (Grantham, 2016; Manabe et al., 2013). On the other hand *esk1*, where xylan acetylation is reduced by 50-60%, the methylation of the GlcA is increased by about 13% (Grantham, 2016; Grantham et al., 2017; Xiong et al., 2013). The opposite effect on the methylation of xylan GlcA in *rwa* and *esk1* mutants is possibly because ESK1 is not responsible for the Ac on 2-*O*-[Me]GlcA-substituted Xyl. Therefore, when ESK1 is inactivated there is more available Ac-CoA channelled to xylan acetylation by TBL32 and TBL33, which leads to a stimulation of the GlcA methylation. The increase of 3-*O*-Ac on 2-*O*-[Me]GlcA substituted Xyl exhibited by *esk1* xylan on the MS analysis (section 6.2.4.1) supports this hypothesis. On the other hand, when only one RWA is active, the levels of Ac-CoA are partially reduced which affect the activity of TBL32 and TBL33, therefore, less

2-*O*-[Me]GlcA-substituted Xyl would be 3-*O*-acetylated, which leads to reduced levels of GlcA methylation. To test this hypothesis, it would be useful to analyse the GlcA methylation levels in other xylan acetylation mutants such as *axy9* or *rwa1/2/3/4*, which should show a stronger GlcA methylation phenotype than the *rwa* triple mutants.

6.3.4 TBL3 is responsible for different xylan Ac modifications to ESK1/TBL28, TBL32/33 and possibly to TBL31

The phylogenetic analysis of ESK1 homologues in angiosperms suggested conserved activities for TBL3 and TBL31. Since each grouped independently with sequences from *Eucalyptus* and *Amborella*, their function is likely to be different.

The promoter swap experiment from section 6.2.2 suggested that TBL3 possibly has a different activity to ESK1 and TBL28. The aim of the promoter swap experiment was to find TBLs with similar activities to ESK1 by expecting a recovery of the *esk1* phenotypes. As discussed before (in section 6.3.2), the restoration of WT traits by the overexpression of ESK1 and TBL28 suggested similar activities. However, the overexpression of TBL3 under the promoter of ESK1 generated instead an aggravation of *esk1* plant growth and caused sterility. In addition, WT lines transformed with the same construct for the overexpression of TBL3, showed WT traits. This suggests that the exacerbated phenotype of *esk1* produced by the overexpression of TBL3 is related to a xylan dysfunction. Therefore, this is evidence suggests that the activity of TBL3 is related to xylan, and that its activity is different to ESK1 and TBL28 since the overexpression of TBL3 did not rescue the growth of *esk1*. None of the other transformed lines showed a clear exacerbation of *esk1* phenotypes, including the overexpression of TBL31.

The PACE analysis using GH11/CE4 (section 6.2.3.5) was used to identify changes on di-substituted Xyl on xylan that are caused when the TBLs are inactivated. In the analysis, *tbl3* xylan showed reproducibly the lack of a single band on PACE. This is indicative of a missing or extra Ac on xylan in this mutant affecting the product outcome of GH11/CE4 digestion. This evidence strongly suggests that TBL3 is responsible for a specific xylan Ac modification. To determine the position of the missing or extra Ac on *tbl3* xylan, it would be necessary to perform an MS analysis of these oligosaccharides.

The activity assay recently reported by Zhong et al. (2017) of purified TBL3 on xylohexaose supports the hypothesis that TBL3 is active on xylan. The assay product detection was carried out using 1-D ^{1}H-NMR, which as discussed above, is not very reliable. Their data suggests that TBL3 is active on xylan, since it produced acetylated products. However, they did not identify the Xyl on the xylohexaose substrate to which the Ac were transferred in the assay. They also did not consider the spontaneous migration of Ac between *O*-2 and *O*-3 (Kabel et al., 2003; Mastihubová and Biely, 2004; Urbanowicz et al., 2014). Since the Ac migration equilibrates with a dominant O-3 linkage, the preference for 3-*O*-acetylation of Xyl by TBL3 reported by Zhong et al. (2017) is possibly an artefact. They also observed a 1-D ^{1}H-NMR peak which they assigned as a ^{1}H on 2,3-di-*O*-acetylated Xyl, however as discussed before, their assignment for X^{aa} is not reliable. Thus, to determine the activity of TBL3 (or any other TBL tested) it would be necessary to analyse the products in the ^{13}C dimension on NMR and by MS experiments.

Since TBL3 showed a specific phenotype when overexpressed in *esk1*, *tbl3* lacked an oligosaccharide on the PACE anaylsis, and TBL31 did not show either phenotype, this suggests that TBL3 and TBL31 have different function and activity. Yuan et al. (2016c) analysed the double mutant *tbl3/31* because their *Arabidopsis* TBL phylogenetic analysis inconsistently suggested that TBL3 and TBL31 may have similar activities. However, they did not find any differences between WT, *tbl3* and *tbl31*. From a group of peaks on 1-D ^{1}H-NMR spectra of xylan from the double mutant *tbl3/31* they interpret a 50% reduction of 3-*O*-Ac and a 25% increase of 2-*O*-Ac on the xylan backbone. Their data suggests that TBL3 is active on xylan, which is supportive of the data presented in this chapter. However, in contrast they proposed similar activities for TBL3 and TBL31, related to 3-*O*-acetylation of xylan Xyl, but their evidence is based on a group of possibly unspecific peaks on 1-D ^{1}H-NMR spectra.

It would be interesting to investigate further the contribution of TBL3 and possibly TBL31 to the acetylation pattern of xylan. The functions of TBL30, TBL34 and TBL35 also remain unresolved. The phylogeny of ESK1 homologues in higher plants presented in this chapter, and the different effects on xylan acetylation found in the *tbl* mutants show clear evidence of that in the process of synthesis of xylan, the acetylation is carried out by a group of acetyltransferases with different regiospecificities. Some of these acetyltransferases belong to the ESK1 closest

paralogues branch, however, other xylan acetyltransferases from other branches of the phylogenetic tree are also possible. The genetic removal of Ac on xylan (in *tbl33/esk1* mutants) produced severely sick plants, which is evidence that Ac are essential for the function of xylan. TBL32 and TBL33 were found to be responsible for 3-*O*-acetylation of 2-*O*-GlcA substituted Xyl, and it would be interesting to investigate whether they are also involved in the acetylation of primary-cell-wall xylan. ESK1 also showed no contribution for the acetylation of primary-cell-wall xylan. ESK1 activity is necessary for the correct [Me]GlcA patterning of secondary cell wall, however, the substitution process of xylan in primary walls may be different. Also, a specific type of xylan may be synthesised in seeds, since GUX2, TBL28 and TBL30 exhibited a differential expression during this stage of the plant. Because GUX2 is normally associated with the co-glucuronidation of xylan with GUX1 in secondary cell walls, but GUX1 did not show expression in seeds, xylan in seeds may have a different structure, modifications and function to the xylan in stems.

Chapter 7: General discussion, conclusions and future work

Cell walls are constituted by several components which interact together, and each of these components play a specific role. Cellulose is the primary and strongest component of the wall, and the cross-linking of the cellulose microfibrils provide support and rigidity to the plant. In eudicot secondary cell walls, xylan is the main hemicellulose, and plays a crucial role coating and cross-linking the cellulose microfibrils (Simmons et al., 2016). The acetylation pattern of xylan has been shown to be crucial for the formation of these interactions (Grantham et al., 2017).

In *Arabidopsis* the degree of acetylation (DA) of xylan has been reported as near 0.60 (Chong et al., 2014). This does not mean that 60% of the backbone Xyl carry an acetyl group, as some carry two. It has been proposed that mostly every other Xyl of the xylan backbone is acetylated (Busse-Wicher et al., 2014; Chong et al., 2014). However, di-acetylated Xyl are only around 6%, which indicates that other acetylation patterns on xylan are possible.

The main aims of this thesis were to describe new acetylation patterns on xylan that are less frequent than the main alternating pattern, and to identify the proteins responsible for the transfer of Ac to each position on xylan. The first aim is discussed in section 7.1 and the second aim in section 7.2.

7.1 Description of xylan acetylation patterns

7.1.1 Acetylation patterns are consistent with the model of compatibility for xylan-cellulose interaction

Bromley et al. (2013) proposed that the activities of GUX1 and GUX2 are narrowed to discrete non-overlapped regions of the xylan backbone. This means that the regions of xylan that are decorated by GUX1 and GUX2 are different. Also, in the absence of GUX1, GUX2 does not extend its activity to where GUX1 normally does, and vice versa. The alteration of xylan acetylation in *esk1* and *axy9* mutants causes a dysregulation of GUX1 and possibly GUX2 (Chapter 6; Grantham, 2016; Grantham et al., 2017) strongly suggests that the compatible and incompatible domains of xylan are determined by specific arrangements of acetylation on the backbone. Therefore,

it is likely that the patterns of acetylation on xylan determine the substrate recognition by GUX1 and GUX2.

Supporting evidence for this hypothesis was presented in Chapter 4, where it was shown that the patterns of acetylation around the GlcA transferred by GUX1 and by GUX2 are different. In regions of xylan decorated by GUX1, mono- or di-acetylated Xyl of the backbone were alternated with unsubstituted Xyl (Figure 4.2). Remarkably, the pattern of acetylated Xyl – non-acetylated Xyl was 'in frame' with the position of the [Me]GlcA modification on the backbone. This implies that the acetylation pattern in xylan regions that are decorated by GUX1 are compatible for interacting with hydrophilic surfaces of cellulose. This evidence strongly supports the hypothesis that GUX1 decorates the compatible domain of xylan. The compatible acetylation pattern of GUX1-decorated xylan is consistent with the main alternating pattern of acetylation that has been previously reported (Figure 1.21; Busse-Wicher et al., 2014; Chong et al., 2014). Since in *Arabidopsis* GUX1 decorates 70% of xylan in secondary cell walls (Bromley et al., 2013), it is likely that the largest acetylation pattern of xylan extends to the whole of the GUX1-decorated region.

The GlcA transferred by GUX2 is surrounded by a distinct (novel) acetylation pattern. This pattern consists on a series of consecutive acetylated Xyl (Figure 4.3). Some non- and di-acetylated Xyl were also found, and their position in the backbone sequence seemed to be stochastic. This acetylation pattern without alternation between substituted with unsubstituted Xyl supports the hypothesis that GUX2 decorates the regions of xylan that do not interact with hydrophilic surfaces of cellulose (Busse-Wicher et al., 2014). In disagreement, molecular-dynamics simulations data shows that oligosaccharides with consecutive MeGlcA-substituted Xyl could interact with hydrophilic surfaces of cellulose (Martínez-Abad et al., 2017). However, in these putative interactions the Xyl and Glc sugars rings are in different planes. Since *gux1* mutant xylan also exhibited a minor production of compatibly-decorated oligosaccharides, it is possible that GUX2 also decorates a small region of the compatible domain of xylan. Alternatively, a minor contamination from primary cell walls was also possible. Supporting evidence for the existence of contiguous acetylated Xyl has been shown by Chong et al. (2014). Their MS analysis of GH10-digested WT xylan showed some minor products with two and three consecutive Xyl that were mono-acetylated. The hydrolysis products structures shown in the

publication had the consecutive acetylated Xyl towards the RE from a [Me]GlcA-substituted Xyl. Since the MS analysis of xylan shown in Chapter 4 in this thesis was carried out using GH30 and not GH10, the portion of xylan analysed may be partially different to that shown by Chong et al. (2014).

Based on the evidence discussed here, a comparative model of the acetylation patterns of xylan in the compatible and incompatible domains is presented in Figure 7.1 (page 251).

The pattern found near [Me]GlcA substitutions in *gux1* mutant xylan is consistent with the clustering of Ac in this domain shown by Grantham (2016). In his doctoral thesis, he observed *gux1* and *gux2* xylan acetylation by using GH10 and GH30 to hydrolyse the xylan, and then he analysed the complete range of oligosaccharides produced, without any enrichment for acidic products nor separation of isomers (e.g. by HILIC). In agreement, he proposed that the incompatible domain is more highly acetylated, based on a higher DA of the oligosaccharides released when digesting *gux1* xylan over *gux2*. He did not determine the specific positions of the Ac respect to the [Me]GlcA substitutions in the hydrolysis products. Supporting evidence for the higher DA on the structures shown in the MS analysis using GH30/GH3 from Chapter 4, and his data shows strong evidence of that near [Me]GlcA substitutions by GUX2 there is a higher DA than near [Me]GlcA by GUX1.

Curiously, $X^{a}X^{a}U^{me,a}X$ originated from *gux2* xylan and does not have a compatible pattern (Figure 4.2E). This oligosaccharide was produced in very low abundance and had three consecutive acetylated-Xyl. Since GUX2 is inactive in this mutant, the MeGlcA on this oligosaccharide is likely a product of GUX1. This indicates that a very low proportion of the GlcA transferred by GUX1 are adjacent to an incompatibly-acetylated region of xylan. Since the activities of GUX1 and GUX2 are each confined to a different region of xylan, it is possible that $X^{a}X^{a}U^{me,a}X$ is produced from the connecting point of the region of xylan decorated by GUX2 (at the NRE) to a region that is decorated by GUX1. Supporting evidence was shown by Bromley et al. (2013), who proposed that GUX1 and GUX2 decorate the same xylan molecules. In the publication they showed that GUX1-decorated xylan and GUX2-decorated xylan are not resolvable by size or charge, which implies that both regions are covalently linked.

7.1.2 Primary-cell-wall xylans are acetylated, and have different substitution patterns that include acetate groups

The function of xylan in the primary cell wall is not yet understood, but curiously, its biosynthesis is performed by a specific and conserved set of enzymes (Mortimer et al., 2015; Ratke et al., 2015). In contrast to secondary cell walls, the lack of xylan in *Arabidopsis* primary cell walls does not seem to affect the plant growth and development, and the proportion of alkali-extractable xylan of primary cell walls is only ~1% compared to the nearly 10% of xylan in woody tissues (Scheller and Ulvskov, 2010; Zablackis et al., 1995). However, maybe a better understanding of the substitutions on this type of xylan could help to elucidate its role.

The structure of xylan from primary walls has been partially characterised in young stems and roots of *Arabidopsis* (Mortimer et al., 2015). It is decorated with GlcA substitutions every six backbone Xyl. In *Arabidopsis* some of these GlcA are unsubstituted, some are methylated, and others are further substituted with an Ara*p* (Mortimer et al., 2015; Peña et al., 2016), but no simultaneous methyl and Ara*p* have been found on the same GlcA. The 1,2-linked Ara*p* on the GlcA was initially reported as a pentopyranose by two groups independently. Mortimer et al. (2015) showed PUX being unique to primary-cell-wall xylan, by showing that it is dependent on the activity of primary-cell-wall xylan synthesis enzymes, IRX9-L, IRX10-L and GUX3. On the other hand, Chong et al. (2015) showed the presence of PUX sidechains on xylan from WT stems, and they also showed a structure containing PUX and an Ac linked to the same Xyl ($P^{a}XX$). Chong et al. (2015) proposed that PUX sidechains are present in primary-cell-wall xylan since they found an enrichment in the secondary-cell-wall-xylan-deficient mutant *irx9*, however they did not show direct evidence to relate PUX to primary-cell-wall xylan. Yet, from the conclusions of both reports taken together it is inferable that primary-cell-wall xylan may be acetylated. Apart from these two publications, no other evidence for acetylation on primary-cell-wall xylan was previously shown.

In this thesis, several structures originated from primary-cell-wall xylan were shown to be acetylated. In Chapter 5, two primary cell wall sources were analysed and compared. Stem primary-cell-wall xylan decorated by GUX3 exhibited Ac decorations at several positions (Figure 5.5). A unique repeating acetylation pattern was found in

this type of xylan, which is illustrated in the model presented in Figure 7.1. 2-*O*-[Me]GlcA- and PUX-substituted Xyl were found 3-*O*-acetylated. The Ac on Xyl carrying a PUX structure is consistent with the structure described by Chong et al. (2014) discussed above. From the [Me]GlcA-substituted Xyl, the second Xyl towards the NRE was found to be predominantly di-acetylated, and the fourth Xyl towards the NRE was found to be either unsubstituted or mono-acetylated. The structures described in this analysis are strong and direct evidence for the presence of acetylation in xylan from primary cell walls. Acetylation patterns of regions of xylan that are glucuronidated by GUX3 have not been previously reported.

The analysis of callus cell walls exhibited a backbone substitution pattern that was to some extent different to the xylan from young stems. The difference in the pattern is not only in the acetylation, but also in Pent modifications on the callus xylan backbone. Most [Me]GlcA on callus xylan were spaced by six backbone residues (Figure 5.6), and Ara*f* substitutions were detected near these [Me]GlcA (Figure 5.7). Callus xylan showed a consistent repeating substitution pattern on the backbone which included Ac and pentosyl substitutions (Figure 5.10). 2-*O*-[Me]GlcA- and 2-*O*-PUX-substituted Xyl were found 3-*O*-acetylated. From the [Me]GlcA-substituted Xyl, the second Xyl towards the NRE was found to be predominantly mono-acetylated, but it was also found carrying a second Ac or a pentosyl substitution. The fourth Xyl from the GlcA towards the NRE was found to be either unsubstituted or mono-arabinosylated. The even number of Xyl of spacing between backbone substitutions implies that the Xyl from the backbone are modified alternately. Since callus and stem primary-wall xylan exhibited substituted Xyl alternated with unsubstituted Xyl, this denotes a similarity between their substitution patterns. Curiously, the positions of the Pent found on the callus-xylan substitution pattern are analogous to Ac found in the pattern from primary-wall xylan from stem.

The pattern of alternating substituted Xyl with unsubstituted Xyl of primary-wall xylan shows a remarkable similarity to the compatible domain of secondary-cell-wall xylan. In both every-other Xyl from the backbone is substituted. Since this pattern is necessary for the normal interaction of xylan with cellulose in a two-fold configuration (Grantham et al., 2017), the pattern on callus and young-stem xylan strongly suggests that the function of xylan in primary cell walls may be to interact with the hydrophilic face of cellulose microfibrils. Supporting evidence for two-fold screw xylan in primary

cell walls is shown in ssNMR spectra of ^{13}C-enriched *Arabidopsis* etiolated hypocotyls by Dick-Perez et al. (2012). At this developmental stage the plant material is enriched for primary cell walls. Their DP-INADEQUATE spectrum shows two unassigned peaks at 82.5-147.5 ppm (*Unk5*) and around 64.0-147.5 ppm (*Unk6*). The peaks with this chemical shift were assigned later in a publication by Simmons et al. (2016) in which a similar experiment was carried out to determine the fold of xylan in the cell wall. The CP-INADEQUATE spectra in the latter publication was acquired from *Arabidopsis* basal stems which are enriched for secondary cell walls, and the cross-peaks corresponding to two-fold xylan Xyl C4 and C5 overlap with the unknown peaks in the former report. This implies that in primary cell walls, two-fold xylan may interact with cellulose microfibrils. In the report by Dick-Perez et al. (2012), the DP-INADEQUATE spectra of depectinated sample with ammonium oxalate showed a reduction of the two possible C4 and C5 cross-peaks from Xyl, and this suggests that some of the xylan may have been removed with the depectination treatment. Therefore, a possible hypothesis for the role of xylan in primary cell walls may be to coat cellulose microfibrils and interact with pectic polysaccharides. Xylan in secondary cell walls are thought to interact with lignin though the GlcA substitutions (Watanabe and Koshijima, 1988), thus, it is possible that the hypothetical xylan-pectin interactions in primary cell walls may be driven by other sidechain structures that are in primary-wall xylan. These structures may be Ara*p* on GlcA (PUX) or Ara*f* substitutions on the backbone. Evidence for the independent interactions arabinogalactan – xylan, and arabinogalactan – pectin were shown by Tan et al. (2013) using NMR experiments. Their data also supports the hypothesis stated here for the interaction xylan – pectin, but through a wall proteic structure (APAP1).

In addition to xylan from callus, possible Ara*f* substitutions on the backbone were found on xylan from mature bottom stems (Chapter 4, Figure 4.3 panels E and Q). If these Pent on the backbone arise from primary cell wall it would have been expected to find increased amounts in stem material with an enrichment for primary cell walls. However, *gux1/2* young stem material (Chapter 5), did not produce any detectable Pent sidechains. It is possible that these xylan Pent side-chains are specific for primary cell walls from mature portions of the stem. Also, the analysis of GUX2-decorated xylan regions (in Chapter 4) is based on the genetic inactivation of GUX1, which removes most of the GlcA on secondary-cell-wall xylan. Therefore, GUX2 may

responsible for the [Me]GlcA substitutions in the vicinity of Pent. However, no Pent sidechains have been previously reported in GUX2-decorated stem xylan (Bromley et al., 2013; Chong et al., 2014; Grantham et al., 2017; Lee et al., 2012; Mortimer et al., 2015, 2010). Alternatively, the *gux1* sample used may have a minor primary cell wall contamination, and since in *gux1* the [Me]GlcA derived from secondary cell walls is reduced, there could be an enrichment for [Me]GlcA derived from primary-cell-wall xylan. In this way, these contaminants may become visible in the mass-spectra. In order to investigate which of the alternatives is correct, it would be useful to carry out the analysis using *gux1/3* mutant xylan, which would have a reduction of the glucuronidated oligosaccharides from primary cell walls. It would also be interesting to investigate further the origin of Ara*f*-substituted primary-wall xylan, since as discussed before, they may drive the link to pectin. Only a single report has previously shown a possible 2-*O*-Ara*f* substitution on the xylan backbone (Tan et al., 2013). This linkage was implied from HSQC and an HMBC NMR spectra of the structure APAP1, which was proposed to link xylan to RG-I and AGPs. In the spectra, 3-*O*-acetylated β-D-Xyl and α-L-Ara*f* were correlated. Since the Ara*f*-substituted Xyl was also shown to be 3-*O*-acetylated, their data is consistent with the structures found in Chapter 5 in primary-wall xylan from callus, in which backbone Xyl were found carrying both Ac and Ara*f*.

Here the substitution pattern of eudicot primary-cell-wall xylan in callus was shown, which has not been previously described. As discussed above, callus xylan has a constant six-Xyl repeating pattern with Ac, [Me]GlcA and Pent modifications (Figure 7.1). A xylan substitution pattern with MeGlcA and Ara*f* has been reported in secondary-cell-wall xylan of the four gymnosperm lineages (Busse-Wicher et al., 2016b). The pattern described showed repeating MeGlcA substitutions every-sixth Xyl. Additionally, cycad, *Ginkgo* and conifer xylan showed Ara*f* located at two and four Xyl of distance from GlcA substitutions towards the NRE (Busse-Wicher et al., 2016b; Martínez-Abad et al., 2017). The xylan of the fourth gymnosperm clade (gnetophytes), which lacks Ara*f*, is acetylated at even spaces (Busse-Wicher et al., 2016b). This indicates that in gymnosperm xylan, substituted Xyl are alternated with non-substituted Xyl. These characteristics of the pattern show remarkable similarity to the ones reported here for callus xylan. Therefore, possibly *Arabidopsis* primary-wall xylan and gymnosperm secondary-wall xylan may have a common xylan ancestor with a similar substitution pattern.

An essential difference between gymnosperm and callus xylan is that in most gymnosperms it lacks acetyl substitutions. Since fern xylan is possibly acetylated (A Li, unpublished data; Haghighat et al., 2016), and gnetophyte xylan is acetylated (Figure 6.2; Busse-Wicher et al., 2016b), other gymnosperms may have lost this function. In agreement with this hypothesis, Pauly et al. (2013) proposed that eudicots may have replaced Ara*f* on xylan with Ac, being the latter more energy-efficient. This is also possible for gnetophytes, which may have conserved xylan acetylation. Supporting data for the gymnosperm loss of function is shown in the phylogenetic analysis of putative acetyltransferases in higher plants (Figure 6.3, page 181), which showed that no gymnosperm representatives had ESK1 homologues. Unfortunately, better-quality data is needed for a more robust phylogenetic analysis of gnetophytes. Yet, investigating xylan acetyltransferases in gnetophytes could produce stronger evidence to support this hypothesis.

The substitution pattern of *Arabidopsis* primary-cell-wall xylan was here shown to be different from the patterns on the compatible and incompatible domains of secondary-cell-wall xylan (Chapter 4). The incompatible domain is different from the others since it exhibited consecutive substituted Xyl (section 7.1.1). On the other hand, the compatible domain showed different positions of di-acetylated Xyl respect to the [Me]GlcA substitutions, while primary-wall xylan normally showed di-acetylated Xyl at two Xyl towards the NRE. In addition, [Me]GlcA on secondary-cell-wall xylan have been shown to be at variable spacings (Chapter 4; Bromley et al., 2013), while in primary walls the spacing is constant (Chapter 5; Mortimer et al., 2015). Since it is possible that a compatible substitution pattern may have appeared first (in gymnosperm/callus xylan ancestor) with a constant six-Xyl repetition pattern, different spacings between Ac and [Me]GlcA may have emerged, originating the two domains of secondary-cell-wall xylan. To test this hypothesis it would be useful to investigate the substitution patterns between primary- and secondary-cell-wall xylan in early-diverging angiosperms and gymnosperms.

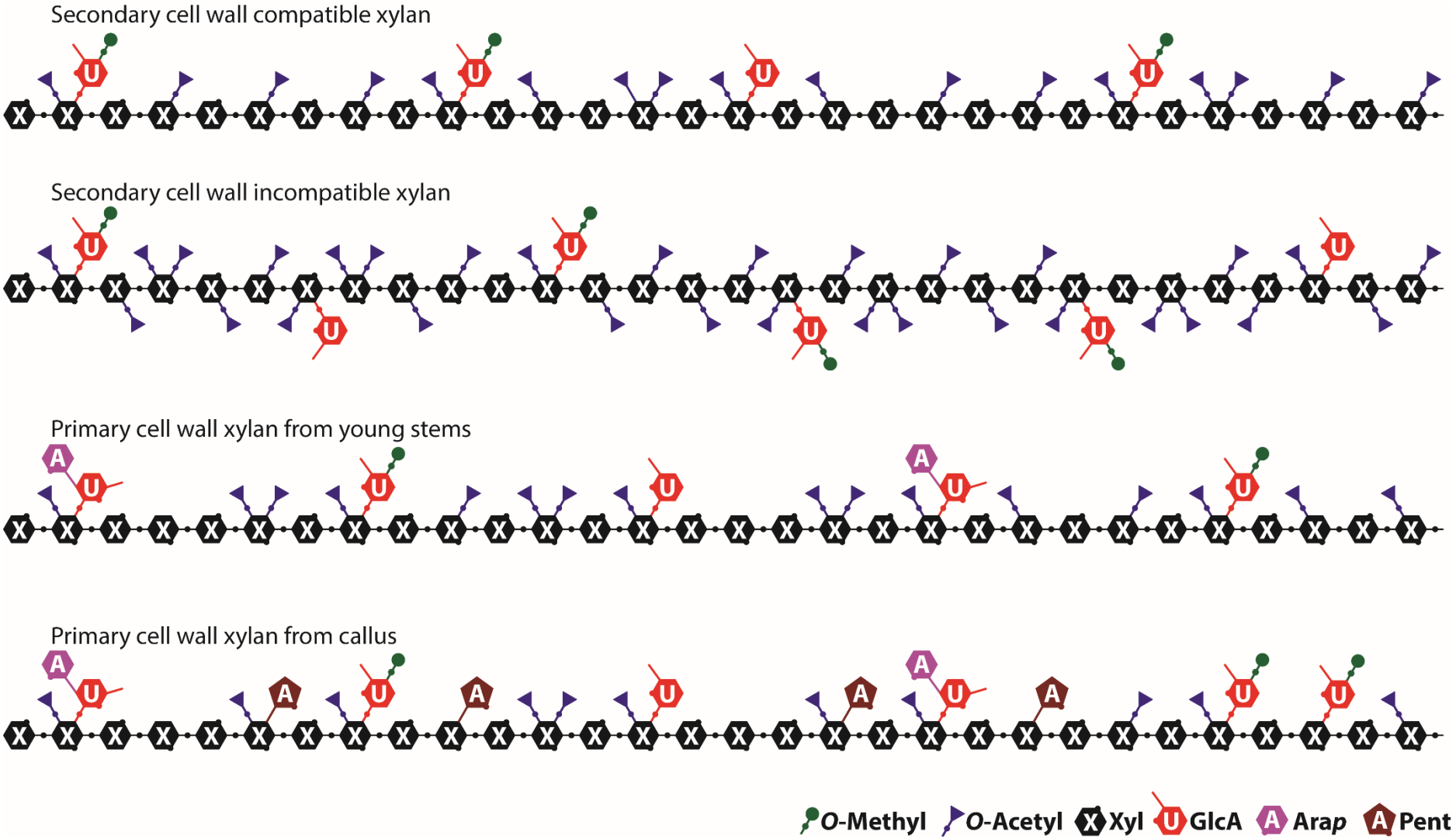


Figure 7.1: Diagram of four differently patterned acetylated xylans found in *Arabidopsis*.
Putative Ara*f* sidechains on callus backbone are drawn as 2-linked Pent for simplicity, but the identity and linkage remains to be determined.

7.2 Acetylation of xylan during biosynthesis

The description of the xylan synthesis pathway has been an aim for more than the last two decades (Bromley et al., 2013; Brown et al., 2009, 2005; Grantham et al., 2017; Urbanowicz et al., 2014, 2012; Wu et al., 2010; Zeng et al., 2016) and the data shown in this thesis is consistent with the current model of xylan biosynthesis. A model of the order of the synthesis in secondary cell walls is presented in Figure 7.2 which is first explained, and then new contributions here proposed are discussed. The reported *in vitro* activity of IRX10, IRX10-L and the XSC is strong evidence of the direction of the xylan synthesis towards the NRE (Jensen et al., 2014; Urbanowicz et al., 2014; Zeng et al., 2016). This direction of synthesis implies that the REO is at the oldest part of the xylan chain, and therefore REO is likely to be a primer for the xylan synthesis. The XSC elongates the β-1,4 Xyl backbone, and shortly after, ESK1 2-*O*-acetylates every other Xyl (Grantham et al., 2017; Urbanowicz et al., 2014). GUX1 and GUX2 then 2-*O*-glucuronidate the xylan (Mortimer et al., 2010; Rennie et al., 2012), each producing specific frequencies of spacing between GlcA modifications in restricted regions of xylan (Bromley et al., 2013). The specific acetylation pattern on the substrate prior to GUX1 and GUX2 activity has not been determined, and whether GUX1 and GUX2 act simultaneously or successively is also unknown. Then, GXM1/2/3 further modifies the GlcA substitutions with a 4-*O*-methyl group (Faria-Blanc, 2014; Grantham, 2016; Li et al., 2013; Yuan et al., 2014). In addition to the synthesis stages that are specific to xylan, RWA1/2/3/4 have been proposed to internalise Ac-CoA to the Golgi cisterna which is donor for Ac modifications (Grantham, 2016; Manabe et al., 2011), and AXY9 has been proposed to form an intermediary between RWA and the acetyltransferases (Schultink et al., 2015). Additional stages are shown in the xylan synthesis model in Figure 7.2 which are proposed in this thesis and are next discussed.

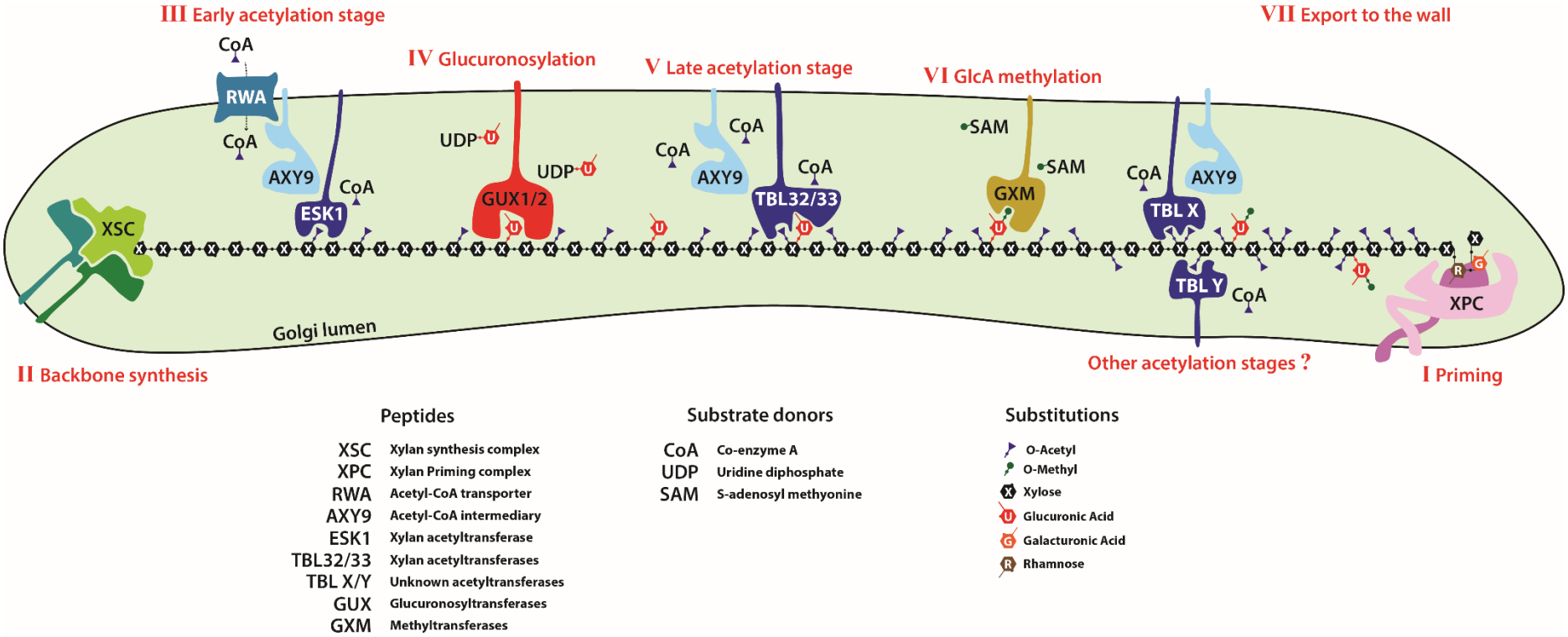


Figure 7.2: Proposed order of synthesis for secondary-cell-wall xylan.
XPC may be composed by IRX7, F8H, IRX8 and PARVUS.
XSC is composed predominantly by IRX9, IRX10 and IRX14.
TBLX is an unknown acetyltransferase activity that transfers a second Ac to mono-acetylated Xyl.
TBLY is an unknown acetyltransferase activity, which is responsible for consecutive acetylated Xyl near GlcA by GUX2.
GXM are GlcA methyltransferases (GXM1/2/3).
RWA are putative Ac-CoA transporters (RWA1/2/3/4).

7.2.1 Xylan acetylation by ESK1

ESK1 is essential for the main pattern of acetylation on xylan (Chapter 6; Grantham, 2016; Grantham et al., 2017; Xiong et al., 2013; Yuan et al., 2013), but the mechanics in which pattern is made is still unknown. Additional xylan acetyltransferases or acetylesterases (Zhang et al., 2017) may also be involved. Since the inactivation of ESK1 affects downstream processes in the synthesis pathway such as the GlcA substitution pattern by GUX1, it is believed that ESK1 acetylates the xylan shortly after the backbone is elongated (Chapter 6; Busse-Wicher et al., 2014; Chong et al., 2014; Grantham, 2016; Grantham et al., 2017; Urbanowicz et al., 2014). Therefore, possibly ESK1 may interact with the XSC. The XSC is composed mainly by IRX9, IRX10 and IRX14 in cells synthesising secondary cell walls, and by IRX9-L, IRX10-L and IRX14 in cells synthesising primary cell walls, although IRX14-L is also involved in the synthesis of both forms of xylan but with a minor degree (Brown et al., 2009, 2007; Mortimer et al., 2015; Wu et al., 2010, 2009). IRX10 and IRX10-L have been shown to have β-1,4-xylosyltransferase activity *in vitro* (Jensen et al., 2014; Urbanowicz et al., 2014), however the functions of IRX9, IRX9-L, IRX14 and IRX14-L is not fully understood. When IRX9, IRX10 or IRX14 is inactivated, the secondary-cell-wall xylan is synthesised by a XSC which is forced to have only the IRX-L homologue in its composition. In Chapter 3, the analysis of xylan on PACE using GH30 revealed that in *irx9*, *irx10* and *irx14* mutants the average spacing between [Me]GlcA is decreased, the compatible pattern generated by GUX1 was altered, and the xylan backbone length was shorter (Figure 3.1 and Figure 3.3). However, when boosting the expression of IRX9-L in *irx9* no xylan phenotypes were visible. This indicates that the phenotypes exhibited by *irx9* xylan are likely due to a deficit of IRX9 function in the XSC, and not because of a different activity of IRX9-L. Therefore, it is possible that no physical interactions are formed between ESK1 and IRX9 during synthesis. Alternatively, XSCs with IRX9 and IRX9-L may both be capable of interacting with ESK1. It is also possible that XSC interacts with ESK1 through IRX10 or IRX14 and not through IRX9.

Since ESK1 acetyltransferase activity is O-2-specific (Urbanowicz et al., 2014), a mechanism is required to prepare the xylan for GUX1 and GUX2 to transfer GlcA to O-2 on Xyl. One hypothesis, proposed by Zhang et al. (2017), is that after xylan is

acetylated by ESK1, acetylesterases may remove some of these Ac. Their evidence shows the presence of Golgi-localised acetylesterases that are specific for xylan. A contrasting hypothesis was proposed by Grantham et al. (2017) in which ESK1 may leave 'gaps' or compete with GUX for substitution of Xyl. To test these hypotheses it would be useful to determine whether ESK1 is sufficient to generate the pattern of acetylation. This could be investigated by determining the position of the Ac transferred *in vitro* by ESK1 to xylooligosaccharides. It would also be informative to investigate the acetylation pattern that each GUX enzyme requires for transferring GlcA. For this it would be useful to determine the position where GlcA is transferred *in vitro* by GUX1/2 onto acetylated/CE4-acetylesterase-treated oligosaccharides.

7.2.1.1 ESK1 activity is necessary for 2,3-di-acetylation

The activity of ESK1 was also here found to be necessary for the generation of di-acetylated Xyl on xylan. The PACE and MS analyses using a combined digestion with GH11 xylanase and CE4 acetylesterase (Chapter 6, Figure 6.14 and Figure 6.16) are methods that showed high sensitivity for the detection of species containing di-acetylated Xyl. WT xylan exhibited di-acetylated-Xyl at several different positions respect to [Me]GlcA substitutions. Contrastingly no di-acetylated Xyl were detected in *esk1* xylan when using both methods. This is strong evidence of that *esk1* xylan lacks di-acetylated Xyl, and therefore it implies that ESK1 is responsible for 2,3-di-acetylated Xyl on xylan in secondary cell walls.

Authors of previous publications have contradicting opinions about ESK1 being responsible for the production of di-acetylated Xyl on xylan. In agreement, Grantham (2016) suggested that ESK1 may be partly responsible for the production of 2,3-di-acetylated Xyl based on 2-D NMR experiments. For WT xylan the peaks related to X23 (X^{aa}, see designations in Figure 6.18) in his ^{13}C-HSQC spectrum were just above the noise levels, and in *esk1* xylan they were absent. However, NMR was only just sensitive enough to detect normal levels of 2,3-di-acetylated Xyl on WT xylan and therefore the lack of a peak which was originally close to the detection limit is not strong evidence for the absence of X^{aa}. In disagreement with these results, Xiong et al. (2013) proposed that ESK1 is not responsible for X^{aa} residues, based on 2-D NMR spectra of *esk1* AIR. However, the low resolution on their spectra, plus the complexity

of the AIR samples may have led them to miss-assign X23 peaks. In four occasions Yuan et al. (2016a, 2016c, 2016b, 2013) showed 1-D ^{1}H-NMR spectra of *esk1* xylan, which they claim shows no reduction in X^{aa} residues. They proposed that ESK1 is not responsible for 2,3-di-acetylated Xyl. However, their data is unreliable, since in contradiction with themselves in two other publications they used *esk1* xylan as 2,3-di-acetylated-Xyl-deficient control (Zhong et al., 2018b, 2017). Therefore, whether ESK1 is responsible for 2,3-di-acetylated Xyl was not truly determined before this thesis work.

How di-acetylated Xyl are synthesised is not yet fully understood. It is possible that the transfer of the second Ac to mono-acetylated Xyl is carried out by ESK1, but for this to occur, the first Ac has to migrate to O-3 to produce a free hydroxyl on C-2 which ESK1 could use as substrate. Alternatively, the transfer of the second Ac may be product of an additional acetyltransferase activity. Since the inactivation of ESK1 also causes an alteration of the pattern of GUX1-mediated [Me]GlcA substitutions on xylan (Chapter 6; Grantham et al., 2017), it is inferred that GUX1 activity happens after xylan acetylation by ESK1. It has not been determined whether the transfer of the second Ac to di-acetylated Xyl, or glucuronidation occurs first. Since di-acetylated Xyl have been found on the secondary-cell-wall xylan in *gux1/2* (Busse-Wicher et al., 2014), it is possible that either di-acetylation occurs after glucuronidation, or that the two processes are independent from each other.

7.2.1.2 ESK1 decorates the compatible and incompatible domains

ESK1 acetylates xylan regions that are downstream glucuronidated by GUX1 and GUX2. In Chapter 4, the acetylation patterns exhibited near [Me]GlcA transferred by GUX1 were different from the environment around [Me]GlcA by GUX2. This is consistent with the hypothesis by Bromley et al. (2013) in which GUX1 and GUX2 recognise specific and independent regions of xylan for the transfer of GlcA. However, 2,3-di-acetylated Xyl were found on both regions of xylan. This is shown in Chapter 4, where the xylan of the mutants *gux1* and *gux2* exhibited di-acetylated Xyl near [Me]GlcA substitutions (Figure 4.2 and Figure 4.3). This implies that 2,3-di-acetylated Xyl exist in regions of secondary-cell-wall xylan that are decorated by GUX1 and by GUX2. Since in *esk1* xylan lacks all di-acetylated Xyl (as discussed above), it is

inferred that during synthesis ESK1 acetylates the regions of xylan decorated by GUX1 and by GUX2. Consistently, Grantham et al. (2017) showed that the [Me]GlcA pattern of xylan produced by GUX1 is altered in *esk1*, and therefore it is evident that part of the Ac on the GUX1-decorated region of xylan is acetylated by ESK1. In contrast, the [Me]GlcA pattern generated by GUX2 in *gux1* mutant has great similarity to the one exhibited in the *esk1* mutant, as the predominant spacings between [Me]GlcA are between five and seven backbone Xyl (Bromley et al., 2013; Grantham et al., 2017). However, the lack of di-acetylated Xyl in *esk1* xylan suggests that ESK1 activity impacts also on the incompatible domain of xylan. It would be interesting to investigate how ESK1 determines the acetylation patterns on the incompatible domain. It would also be interesting to determine whether consecutive mono-acetylated Xyl exist near [Me]GlcA in *esk1* xylan, since it is an acetylation pattern that is characteristic of the incompatible domain of xylan (Chapter 4).

7.2.2 TBL32 and TBL33 acetylate the compatible and incompatible domains

A second stage of xylan acetylation is carried out by TBL32 and TBL33 after the glucuronidation by GUX1/2. In Chapter 6, *tbl32* and *tbl33* mutants exhibited a partial loss of Ac on $U^{[me],a}$ residues, and in *tbl32/33* acetylated $U^{[me],a}$ residues were barely undetectable. Therefore, it was concluded that TBL32 and TBL33 are xylan acetyltransferases. Since the [Me]GlcA levels and spacing pattern in *tbl32/33* xylan are unaltered, it implies that GUX1 and GUX2 do not require the specific Ac by TBL32/33 for their activity. On the other hand, during the course of work in this thesis, Zhong et al. (2017) showed that TBL32 and TBL33 seem to require a GlcA modification in the acceptor for their activity, which is consistent with the order of synthesis here proposed. Therefore, it is implied that the acetylation of xylan by TBL32 and TBL33 occurs downstream to xylan glucuronidation by GUX1 and GUX2. The position of TBL32/33-mediated Ac was confirmed in the ^{13}C-HSQC (Figure 6.18), in which the only Ac sensitive to the activity of TBL32 and TBL33 was O-3-linked to 2-*O*-[Me]GlcA-substituted Xyl.

The novel acetyltransferases identified TBL32 and TBL33 are essential for xylan function. The combined inactivation of TBL33 and ESK1 generated a severely sick plants, sterile and with dwarfed growth (Figure 6.10). This phenotype is similar to that

exhibited by *irx9/9-l*, *irx10/10-l* and *irx14/14-l* which have no xylan (Brown et al., 2009, 2007; Wu et al., 2010, 2009). ESK1, TBL33 are possibly the two main xylan acetyltransferases, and since partial deficits in acetylation in *esk1* or *rwa* mutants produces mild to strong growth phenotypes and collapsed xylem vessels (Lefebvre et al., 2011; Manabe et al., 2013), it was expected that the co-inactivation of ESK1 and another xylan acetyltransferase would lead to further non-functionality of xylan. Also, the lack of TBL33 and TBL32 was here shown to have an impact on the downstream methylation of the GlcA. The DASH analysis of *tbl32/33* xylan showed a significant reduction of GlcA methylation by about 50%. Xylan GlcA methylation has been proposed to facilitate the formation of the cross-link of xylan to lignin in eudicots (Imamura et al., 1994; Watanabe and Koshijima, 1988). Since the angiosperm conserved TBL32/33 function boosts the methylation levels of GlcA, the Ac they mediate might also influence the cross-linking process (Terrett and Dupree, 2019). The NMR spectra of *tbl32* and *tbl33* (Figure 6.18) showed shifted peaks for XG3 (-CH_3) and XG3 (H3C3 and H4C4) for an unknown reason. This shift may be suggestive of a different environment around the xylan [Me]GlcA-substituted Xyl when they carry a 3-*O*-Ac.

The model of xylan biosynthesis presented in Figure 7.2 shows all steps discussed here, however, several steps are still yet to be resolved.

7.3 Unresolved issues

7.3.1 Description of new substitution patterns on xylan

The acetylation patterns described in this thesis are only narrowed to what was observed in stems and callus. However, studies on seed mucilage have shown particular side-chain substitutions on xylan (Voiniciuc et al., 2015). These side-chains were shown to be required for cross-linking to pectin (Ralet et al., 2016). It is possible that xylan in other tissues or cell types may have additional roles, which still remain to be elucidated.

For the description of the acetylation patterns of xylan, GH30/GH3 were used to hydrolyse the xylan, and then the products were analysed by PACE and mass-spectrometry. GH10 products were analysed by PACE but the MS analysis was not

possible since the required optimisation of the HILIC separation was not achieved. Because of the activity of GH30, the description of its products provides information of the decorations present towards the NRE of the [Me]GlcA substitutions. Since [Me]GlcA is not required to be in a specific subsite for GH10 hydrolysis, the products contain more information than what is observable using GH30. In addition, it is likely that GH10 is capable of accessing regions on xylan that are inaccessible for GH30.

7.3.2 Xylan recognition sites by GUX1, GUX2 and GUX3

The acetylation contexts of GUX1, GUX2 and GUX3-decorated regions of xylan are different (Chapter 4 and Chapter 5). However, it is possible that the acetylation context on xylan during the synthesis differs to the final product *in muro*. It has been recently identified that GUX1 and GUX2 are more active on partially de-acetylated substrates (Łyczakowski, 2018). In order to determine the acetylation context that GUX1, GUX2 and GUX3 recognise, it would be interesting to assay their activity *in vitro* on xylan with different degrees of acetylation. The MS analysis of their activities would provide great contribution towards understanding the acetylation context in which they perform their activities during synthesis.

As discussed above, ESK1-mediated Ac are required for the correct regulation of GUX1 and GUX2 to transfer GlcA to secondary-cell-wall xylan with the correct spacing pattern. However, it remains unknown whether Ac of primary-cell-wall xylan is necessary for the GUX3-mediated [Me]GlcA spacing pattern. To investigate the impact of the loss of primary-cell-wall xylan acetylation, on downstream xylan substitution processes, it would be interesting to investigate the [Me]GlcA pattern of primary-wall xylan in the xylan-acetylation-deficient mutant *axy9*. If the inactivation of AXY9 affects the spacing of [Me]GlcA on primary-cell-wall xylan, it would imply that the activity of GUX3 may depend on an upstream xylan-acetylation process.

7.3.3 Other possible xylan acetyltransferases

The identified TBL activities that showed to be responsible for the transfer of Ac to specific positions and patterns on xylan (Chapter 6) do not cover the entire range of Ac arrangements that were found on primary- and secondary-cell-wall xylan (Chapter 4 and Chapter 5). It was also not determined whether any of the identified activities is

responsible for the acetylation of primary-cell-wall xylan. None of the *tbl* mutants analysed showed a clear alteration of primary-cell-wall-xylan PUX spacing pattern (Figure 6.12), and the acetylation patterns on primary-cell-wall xylan in *esk1/tbl28*, *tbl32/33* and *tbl3* were not directly analysed. Since the phylogenetic analysis of ESK1 closest homologues in higher plants showed that five or six activities that are conserved in eudicots (Figure 6.3), it is likely that TBL30, TBL31, TBL34 and TBL35 may have a different regiospecificity. Perhaps this was not yet revealed because the methods used were not appropriate to investigate Ac on xylan that are dependent on their activities. However, Zhong et al. (2017) suggested that these proteins may be active on xylan. Therefore, possibly some of these ESK1 high-similarity paralogues may be responsible for the acetylation of minor regions of xylan, or xylan in other tissues. Since there was no evidence of that TBL family members outside the ESK1-closest-homologues clade may be active on xylan, they were not investigated here. No ferns or *Gnetum* ESK1 homologs grouped in ESK1 closest-homologs clade, but since *Gnetum* xylan is acetylated (Busse-Wicher et al., 2016b) and possibly fern xylan as well (A Li, unpublished; Haghighat et al., 2016a), it is very likely that some TBLs from other parts of the tree may also be active on xylan.

TBL3 was determined to be responsible for a specific xylan modification which is likely to be an Ac related to the acetylation pattern, but the positions of this Ac was not identified. *tbl3* mutant xylan showed a single band difference on PACE when digested with a xylanase (Figure 6.14). Additionally, all *esk1* lines overexpressing TBL3 under the promoter of ESK1 showed a severe exacerbation of the dwarfed growth and caused sterility (Figure 6.10). In these overexpressor lines, TBL3 may transfer Ac onto the xylan backbone in places or positions that causes xylan to be less functional than in *esk1*. The data from the TBL3 overexpressing lines and from the PACE analysis is strong evidence of that TBL3 is active on xylan, and it would be interesting to investigate its activity.

In conclusion, this study has allowed for significant steps to be taken towards the characterization of the structure and biosynthesis of xylan in the eudicot model *Arabidopsis*. Novel substitution patterns of primary- and secondary-cell-wall xylan were described in detail, which contributes to the model of xylan-cellulose interaction. The role of primary-cell-wall xylan and the importance of its even pattern of substitutions remain to be elucidated. Also, TBL32 and TBL33 were shown to

participate on the glucuronoxylan acetylation process. Since other possible xylan acetyltransferases were also identified, their specific role in xylan substitution remains to be determined. The new knowledge contributed by this thesis is directly useful for plant-cell-wall structure engineering and synthetic biology projects that are currently being carried out in the Dupree lab.

Chapter 8: Bibliography

Akoh, C.C., Lee, G.C., Liaw, Y.C., Huang, T.H., Shaw, J.F., 2004. GDSL family of serine esterases/lipases. Prog. Lipid Res. doi:10.1016/j.plipres.2004.09.002

Anantharaman, V., Aravind, L., 2010. Novel eukaryotic enzymes modifying cell-surface biopolymers. Biol. Direct 5. doi:10.1186/1745-6150-5-1

Anders, N., Wilkinson, M.D., Lovegrove, A., Freeman, J., Tryfona, T., Pellny, T.K., Weimar, T., Mortimer, J.C., Stott, K., Baker, J.M., Defoin-Platel, M., Shewry, P.R., Dupree, P., Mitchell, R.A.C., 2012. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. Proc. Natl. Acad. Sci. 109, 989–993. doi:10.1073/pnas.1115858109

Aspinall, G.O., 1980. Chemistry of Cell Wall Polysaccharides, Carbohydrates: Structure and Function. Academic Press. doi:10.1016/B978-0-12-675403-2.50018-1

Atalla, R.H., Nagel, S.C., 1984. Cellulose: Its regeneration in the native lattice. Science (80-.). 185, 522–523. doi:10.1126/science.185.4150.522

Bar-On, Y.M., Phillips, R., Milo, R., 2018. The biomass distribution on Earth. Proc. Natl. Acad. Sci. 115, 6506–6511. doi:10.1073/pnas.1711842115

Barton, C.J., Tailford, L.E., Welchman, H., Zhang, Z., Gilbert, H.J., Dupree, P., Goubet, F., 2006. Enzymatic fingerprinting of Arabidopsis pectic polysaccharides using polysaccharide analysis by carbohydrate gel electrophoresis (PACE). Planta 224, 163–174. doi:10.1007/s00425-005-0185-9

Bassel, G.W., Fung, P., Chow, T. -f. F., Foong, J.A., Provart, N.J., Cutler, S.R., 2008. Elucidating the Germination Transcriptional Program Using Small Molecules. PLANT Physiol. 147, 143–155. doi:10.1104/pp.107.110841

Bensussan, M., Lefebvre, V., Ducamp, A., Trouverie, J., Gineau, E., Fortabat, M.-N., Guillebaux, A., Baldy, A., Naquin, D., Herbette, S., Lapierre, C., Mouille, G., Horlow, C., Durand-Tardif, M., 2015. Suppression of Dwarf and *irregular xylem* Phenotypes Generates Low-Acetylated Biomass Lines in Arabidopsis. Plant Physiol. 168, 452–463. doi:10.1104/pp.15.00122

Biely, P., Cziszárová, M., Uhliariková, I., Agger, J.W., Li, X.L., Eijsink, V.G.H., Westereng, B., 2013. Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. Biochim. Biophys. Acta - Gen. Subj. 1830, 5075–5086. doi:10.1016/j.bbagen.2013.07.018

Biely, P., De Vries, R.P., Vršanská, M., Visser, J., 2000. Inverting character of α-glucuronidase A from Aspergillus tubingensis. Biochim. Biophys. Acta - Gen. Subj. 1474, 360–364. doi:10.1016/S0304-4165(00)00029-5

Biely, P., MacKenzie, C.R., Puls, J., Schneider, H., 1986. Cooperativity of Esterases and Xylanases in the Enzymatic Degradation of Acetyl Xylan. Bio/Technology 4, 731–733. doi:10.1038/nbt0886-731

Biely, P., Vršanská, M., Tenkanen, M., Kluepfel, D., 1997. Endo-β-1,4-xylanase families: Differences in catalytic properties. J. Biotechnol. 57, 151–166. doi:10.1016/S0168-1656(97)00096-5

Bischoff, V., Nita, S., Neumetzler, L., Schindelasch, D., Urbain, A., Eshed, R., Persson, S., Delmer, D., Scheible, W.-R., 2010a. TRICHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose Biosynthesis in Arabidopsis. Plant Physiol. 153, 590–602. doi:10.1104/pp.110.153320

Bischoff, V., Selbig, J., Scheible, W.-R., 2010b. Involvement of TBL/DUF231 proteins into cell wall biology. Plant Signal. Behav. 5, 1057–1059. doi:10.4161/psb.5.8.12414

Boerjan, W., Ralph, J., Baucher, M., 2003. Lignin Biosynthesis. Annu. Rev. Plant Biol. 54, 519–546. doi:10.1146/annurev.arplant.54.031902.134938

Bonawitz, N.D., Chapple, C., 2010. The Genetics of Lignin Biosynthesis: Connecting Genotype to Phenotype. Annu. Rev. Genet. 44, 337–363. doi:10.1146/annurev-genet-102209-163508

Bouchabke-Coussa, O., Quashie, M.-L., Seoane-, J., Fortabat, M.-N., Gery, C., Yu, A., Linderme, D., Trouverie, J., Granier, F., Téoulé, E., Durand-tardif,

M., 2008. acclimation and salt tolerance. BMC Plant Biol. 27, 1–27. doi:10.1186/1471-2229-8-125

Bromley, J.R., Busse-Wicher, M., Tryfona, T., Mortimer, J.C., Zhang, Z., Brown, D.M., Dupree, P., 2013. GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. Plant J. 74, 423–434. doi:10.1111/tpj.12135

Brown, D.M., Goubet, F., Wong, V.W., Goodacre, R., Stephens, E., Dupree, P., Turner, S.R., 2007. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J. 52, 1154–1168. doi:10.1111/j.1365-313X.2007.03307.x

Brown, D.M., Wightman, R., Zhang, Z., Gomez, L.D., Atanassov, I., Bukowski, J.P., Tryfona, T., McQueen-Mason, S.J., Dupree, P., Turner, S.R., 2011. Arabidopsis genes IRREGULAR XYLEM (IRX15) and IRX15L encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. Plant J. 66, 401–413. doi:10.1111/j.1365-313X.2011.04501.x

Brown, D.M., Zeef, L.A.H., Ellis, J., Goodacre, R., Turner, S.R., 2005. Identification of Novel Genes in Arabidopsis Involved in Secondary Cell Wall Formation Using Expression Profiling and Reverse Genetics. Plant Cell 17, 2281–2295. doi:10.1105/tpc.105.031542.major

Brown, D.M., Zhang, Z., Stephens, E., Dupree, P., Turner, S.R., 2009. Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. Plant J. 57, 732–746. doi:10.1111/j.1365-313X.2008.03729.x

Buckeridge, M.S., Rayon, C., Urbanowicz, B., Tiné, M.A.S., Carpita, N.C., 2004. Mixed Linkage (1→3),(1→4)-β-D-Glucans of Grasses. Cereal Chem. doi:10.1094/CCHEM.2004.81.1.115

Buliga, G.S., Brant, D.A., Fincher, G.B., 1986. The sequence statistics and solution conformation of a barley (1→3, 1→)-β-d-glucan. Carbohydr. Res. 157, 139–156. doi:10.1016/0008-6215(86)85065-0

Burton, R.A., Gidley, M.J., Fincher, G.B., 2010. Heterogeneity in the chemistry, structure and function of plant cell walls. Nat. Chem. Biol. doi:10.1038/nchembio.439

Burton, R.A., Wilson, S.M., Hrmova, M., Harvey, A.J., Shirley, N.J., Medhurst, A., Stone, B.A., Newbigin, E.J., Bacic, A., Fincher, G.B., 2006. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-glucans. Science (80-.). 311, 1940–1942. doi:10.1126/science.1122975

Busse-Wicher, M., Gomes, T.C.F., Tryfona, T., Nikolovski, N., Stott, K., Grantham, N.J., Bolam, D.N., Skaf, M.S., Dupree, P., 2014. The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana. Plant J. 79, 492–506. doi:10.1111/tpj.12575

Busse-Wicher, M., Grantham, N.J., Lyczakowski, J.J., Nikolovski, N., Dupree, P., 2016a. Xylan decoration patterns and the plant secondary cell wall molecular architecture. Biochem. Soc. Trans. 44, 74–78. doi:10.1042/BST20150183

Busse-Wicher, M., Li, A., Silveira, R.L., Pereira, C.S., Tryfona, T., Gomes, T.C.F., Skaf, M.S., Dupree, P., 2016b. Evolution of xylan substitution patterns in gymnosperms and angiosperms: implications for xylan interaction with cellulose. Plant Physiol. 171, pp.00539.2016. doi:10.1104/pp.16.00539

Caffall, K.H., Mohnen, D., 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr. Res. 344, 1879–1900. doi:10.1016/j.carres.2009.05.021

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 37, 233–238. doi:10.1093/nar/gkn663

Carpita, N.C., 1996. STRUCTURE AND BIOGENESIS OF THE CELL WALLS OF GRASSES. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47, 445–476. doi:10.1146/annurev.arplant.47.1.445

Carpita, N.C., Gibeaut, D.M., 1993. Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3, 1–30. doi:10.1111/j.1365-313X.1993.tb00007.x

Chesson, A., Gordon, A.H., Lomax, J.A., 1983. Substituent groups linked by alkali- labile bonds to arabinose and xylose residues of legume, grass and cereal

straw cell walls and their fate during digestion by rumen microorganisms. J. Sci. Food Agric. 34, 1330–1340. doi:10.1002/jsfa.2740341204

Chiniquy, D., Sharma, V., Schultink, A., Baidoo, E.E., Rautengarten, C., Cheng, K., Carroll, A., Ulvskov, P., Harholt, J., Keasling, J.D., Pauly, M., Scheller, H.V., Ronald, P.C., 2012. XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. Proc. Natl. Acad. Sci. U. S. A. 109, 17117–22. doi:10.1073/pnas.1202079109

Chong, S.-L., Koutaniemi, S., Juvonen, M., Derba-Maceluch, M., Mellerowicz, E.J., Tenkanen, M., 2015. Glucuronic acid in Arabidopsis thaliana xylans carries a novel pentose substituent. Int. J. Biol. Macromol. 79, 807–812. doi:10.1016/j.ijbiomac.2015.05.056

Chong, S.-L., Virkki, L., Maaheimo, H., Juvonen, M., Derba-Maceluch, M., Koutaniemi, S., Roach, M., Sundberg, B., Tuomainen, P., Mellerowicz, E.J., Tenkanen, M., 2014. O-Acetylation of glucuronoxylan in arabidopsis thaliana wild type and its change in xylan biosynthesis mutants. Glycobiology 24, 494–506. doi:10.1093/glycob/cwu017

Coenen, G.J., Bakx, E.J., Verhoef, R.P., Schols, H.A., Voragen, A.G.J., 2007. Identification of the connecting linkage between homo- or xylogalacturonan and rhamnogalacturonan type I. Carbohydr. Polym. 70, 224–235. doi:10.1016/j.carbpol.2007.04.007

Cornuault, V., Buffetto, F., Rydahl, M.G., Marcus, S.E., Torode, T.A., Xue, J., Crépeau, M.J., Faria-Blanc, N., Willats, W.G.T., Dupree, P., Ralet, M.C., Knox, J.P., 2015. Monoclonal antibodies indicate low-abundance links between heteroxylan and other glycans of plant cell walls. Planta 242, 1321–1334. doi:10.1007/s00425-015-2375-4

Cosgrove, D.J., 2014. Re-constructing our models of cellulose and primary cell wall assembly. Curr. Opin. Plant Biol. 22, 122–131. doi:10.1016/j.pbi.2014.11.001

Cosgrove, D.J., 2005. Growth of the plant cell wall. Nat. Rev. Mol. Cell Biol. doi:10.1038/nrm1746

Darvill, J.E., McNeil, M., Darvill, A.G., Albersheim, P., 1980. Structure of Plant Cell Walls: XI. Glucuronoarabinoxylan, a second hemicellulose in the primary cell walls of suspension-cultured sycamore cells. Plant Physiol. 66, 1135–9. doi:10.1104/pp.66.6.1135

Davies, G.J., Wilson, K.S., Henrissat, B., 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. 321 (Pt 2, 557–559. doi:10.1007/s007920050009

Del Río, J.C., Marques, G., Rencoret, J., Martínez, Á.T., Gutiérrez, A., 2007. Occurrence of naturally acetylated lignin units. J. Agric. Food Chem. 55, 5461–5468. doi:10.1021/jf0705264

Derba-Maceluch, M., Awano, T., Takahashi, J., Lucenius, J., Ratke, C., Kontro, I., Busse-Wicher, M., Kosik, O., Tanaka, R., Winzéll, A., Kallas, Å., Leśniewska, J., Berthold, F., Immerzeel, P., Teeri, T.T., Ezcurra, I., Dupree, P., Serimaa, R., Mellerowicz, E.J., 2015. Suppression of xylan endotransglycosylase PtxtXyn10A affects cellulose microfibril angle in secondary wall in aspen wood. New Phytol. 205, 666–681. doi:10.1111/nph.13099

Desprez, T., Juraniec, M., Crowell, E.F., Jouy, H., Pochylova, Z., Parcy, F., Hofte, H., Gonneau, M., Vernhettes, S., 2007. Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. Proc. Natl. Acad. Sci. 104, 15572–15577. doi:10.1073/pnas.0706569104

Dick-Perez, M., Wang, T., Salazar, A., Zabotina, O.A., Hong, M., 2012. Multidimensional solid-state NMR studies of the structure and dynamics of pectic polysaccharides in uniformly13C-labeled Arabidopsis primary cell walls. Magn. Reson. Chem. 50, 539–550. doi:10.1002/mrc.3836

Dickison, W.C., 2000. Integrative Plant Anatomy, Integrative Plant Anatomy. Academic Press. doi:10.1016/B978-012215170-5/50003-0

Dodd, D., Cann, I.K.O., 2009. Enzymatic deconstruction of xylan for biofuel production. GCB Bioenergy 1, 2–17. doi:10.1111/j.1757-1707.2009.01004.x

Domon, B., Costello, C.E., 1988. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. Glycoconj. J. 5, 397–409. doi:10.1007/BF01049915

Dupree, R., Simmons, T.J., Mortimer, J.C., Patel, D., Iuga,

D., Brown, S.P., Dupree, P., 2015. Probing the Molecular Architecture of *Arabidopsis thaliana* Secondary Cell Walls Using Two- and Three-Dimensional ^{13} C Solid State Nuclear Magnetic Resonance Spectroscopy. Biochemistry 54, 2335–2345. doi:10.1021/bi501552k

Ebringerová, A., Heinze, T., 2000. Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. Macromol. Rapid Commun. 21, 542–556. doi:10.1002/1521-3927(20000601)21:9<542::AID-MARC542>3.3.CO;2-Z

Enebro, J., Karlsson, S., 2006. Improved matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry of carboxymethyl cellulose. Rapid Commun. Mass Spectrom. 20, 3693–3698. doi:Doi 10.1002/Rcm.2786

Evtuguin, D. V., Tomás, J.L., Silva, A.M.S., Neto, C.P., 2003. Characterization of an acetylated heteroxylan from Eucalyptus globulus Labill. Carbohydr. Res. 338, 597–604. doi:10.1016/S0008-6215(02)00529-3

Faria-Blanc, N., 2014. Xylan Methylation and Synthesis in Arabidopsis. University of Cambridge.

Faria-Blanc, N., Mortimer, J.C., Dupree, P., 2018. A Transcriptomic Analysis of Xylan Mutants Does Not Support the Existence of a Secondary Cell Wall Integrity System in Arabidopsis. Front. Plant Sci. 9, 384. doi:10.3389/fpls.2018.00384

Feijão, C. de V.G.P., 2016. Characterization of grass xylan structure and biosynthesis by. University of Cambridge.

Fernandes, A.N., Thomas, L.H., Altaner, C.M., Callow, P., Forsyth, V.T., Apperley, D.C., Kennedy, C.J., Jarvis, M.C., 2011. Nanostructure of cellulose microfibrils in spruce wood. Proc. Natl. Acad. Sci. 108, E1195–E1203. doi:10.1073/pnas.1108942108

Food and Agriculture Organization - FAO, 2018. Forest Products 2016 yearbook.

French, A.D., Johnson, G.P., 2009. Cellulose and the twofold screw axis: Modeling and experimental arguments. Cellulose 16, 959–973. doi:10.1007/s10570-009-9347-4

Fry, S.C., 2011. Cell Wall Polysaccharide Composition and Covalent Crosslinking, in: Annual Plant Reviews. pp. 1–42. doi:10.1002/9781444391015.ch1

Fry, S.C., 1986. Cross-Linking of Matrix Polymers in the Growing Cell Walls of Angiosperms. Annu. Rev. Plant Physiol. 37, 165–186. doi:10.1146/annurev.pp.37.060186.001121

Fry, S.C., Miller, J.C., 1989. Toward a working model of the growing plant cell wall. Phenolic cross-linking reactions in the primary cell walls of dicotyledons, in: Plant Cell Wall Polymers. pp. 33–46. doi:10.1021/bk-1989-0399.ch003

Fry, S.C., Nesselrode, B.H.W.A., Miller, J.G., Mewburn, B.R., 2008. Mixed-linkage (1→3,1→4)-β-D-glucan is a major hemicellulose of Equisetum (horsetail) cell walls. New Phytol. 179, 104–115. doi:10.1111/j.1469-8137.2008.02435.x

Fry, S.C., York, W.S., Albersheim, P., Darvill, A.G., Hayashi, T., Joseleau, J. -P, Kato, Y., Lorences, E.P., Maclachlan, G.A., McNeil, M., Mort, A.J., Grant Reid, J.S., Seitz, H.U., Selvendran, R.R., Voragen, A.G.J., White, A.R., 1993. An unambiguous nomenclature for xyloglucan-derived oligosaccharides. Physiol. Plant. 89, 1–3. doi:10.1111/j.1399-3054.1993.tb01778.x

Gille, S., de Souza, A., Xiong, G., Benz, M., Cheng, K., Schultink, A., Reca, I.-B., Pauly, M., 2011. O-acetylation of Arabidopsis hemicellulose xyloglucan requires AXY4 or AXY4L, proteins with a TBL and DUF231 domain. Plant Cell 23, 4041–53. doi:10.1105/tpc.111.091728

Gille, S., Pauly, M., 2012. O-Acetylation of Plant Cell Wall Polysaccharides. Front. Plant Sci. 3, 12. doi:10.3389/fpls.2012.00012

Gonçalves, V.M.F., Evtuguin, D. V., Domingues, M.R.M., 2008. Structural characterization of the acetylated heteroxylan from the natural hybrid Paulownia elongata/Paulownia fortunei. Carbohydr. Res. 343, 256–266. doi:10.1016/j.carres.2007.11.002

Gonneau, M., Desprez, T., Guillot, A., Vernhettes, S., Höfte, H., 2014. Catalytic subunit stoichiometry within the cellulose synthase complex. Plant Physiol. 166, 1709–12. doi:10.1104/pp.114.250159

Goubet, F., Jackson, P., Deery, M.J., Dupree, P., 2002.

Polysaccharide Analysis Using Carbohydrate Gel Electrophoresis: A Method to Study Plant Cell Wall Polysaccharides and Polysaccharide Hydrolases. Anal. Biochem. 300, 53–68. doi:10.1006/abio.2001.5444

Grantham, N.J., 2016. The role of acetylation in [methyl] glucuronic acid patterning and synthesis of Arabidopsis xylan. University of Cambridge.

Grantham, N.J., Wurman-Rodrich, J., Terrett, O.M., Lyczakowski, J.J., Stott, K., Iuga, D., Simmons, T.J., Durand-Tardif, M., Brown, S.P., Dupree, R., Busse-Wicher, M., Dupree, P., 2017. An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. Nat. Plants 3, 859–865. doi:10.1038/s41477-017-0030-8

Guo, Q., Cui, S.W., Wang, Q., Christopher Young, J., 2008. Fractionation and physicochemical characterization of psyllium gum. Carbohydr. Polym. 73, 35–43. doi:10.1016/j.carbpol.2007.11.001

Hägglund, E., Lindberg, B., McPherson, J., 1956. Dimethylsulphoxide, a Solvent for Hemicelluloses. Acta Chem. Scand. 10, 1160–1164.

Haghighat, M., Teng, Q., Zhong, R., Ye, Z.-H.H., 2016. Evolutionary conservation of xylan biosynthetic genes in selaginella moellendorffii and physcomitrella patens. Plant Cell Physiol. 57, 1707–1719. doi:10.1093/pcp/pcw096

Hall, B.G., 2013. Building phylogenetic trees from molecular data with MEGA. Mol. Biol. Evol. 30, 1229–1235. doi:10.1093/molbev/mst012

Handford, M.G., Baldwin, T.C., Goubet, F., Prime, T.A., Miles, J., Yu, X., Dupree, P., 2003. Localisation and characterisation of cell wall mannan polysaccharides in Arabidopsis thaliana. Planta 218, 27–36. doi:10.1007/s00425-003-1073-9

Hao, Z., Mohnen, D., 2014. A review of xylan and lignin biosynthesis: Foundation for studying Arabidopsis *irregular xylem* mutants with pleiotropic phenotypes. Crit. Rev. Biochem. Mol. Biol. 49, 212–241. doi:10.3109/10409238.2014.889651

Harholt, J., Suttangkakul, A., Vibe Scheller, H., 2010. Biosynthesis of pectin. Plant Physiol. 153, 384–95. doi:10.1104/pp.110.156588

Harvey, D., 2000. Modern Analytical Chemistry. doi:10.1136/jcp.30.1.93-a

Hill, J.L., Hammudi, M.B., Tien, M., 2014. The Arabidopsis Cellulose Synthase Complex: A Proposed Hexamer of CESA Trimers in an Equimolar Stoichiometry. Plant Cell Online 26, 4834–4842. doi:10.1105/tpc.114.131193

Hinz, S.W.A., Verhoef, R., Schols, H.A., Vincken, J.P., Voragen, A.G.J., 2005. Type I arabinogalactan contains β-D-Galp-(1→3)-β-D-Galp structural elements. Carbohydr. Res. 340, 2135–2143. doi:10.1016/j.carres.2005.07.003

Imamura, T., Watanabe, T., Kuwahara, M., Koshijima, T., 1994. Ester linkages between lignin and glucuronic acid in lignin-carbohydrate complexes from Fagus crenata. Phytochemistry 37, 1165–1173. doi:10.1016/S0031-9422(00)89551-5

Ishii, T., 1997. O-acetylated oligosaccharides from pectins of potato tuber cell walls. Plant Physiol. 113, 1265–72. doi:10.1104/pp.113.4.1265

Ishii, T., 1995. Isolation and Characterization of Acetylated Rhamnogalacturonan Oligomers Liberated From Bamboo Shoot Cell-Walls By Driselase. Mokuzai Gakkaishi 41, 561–572.

Ivakov, A., Persson, S., 2012. Plant Cell Walls, in: ELS. doi:10.1002/9780470015902.a0001682.pub2

Janbon, G., Himmelreich, U., Moyrand, F., Improvisi, L., Dromer, F., 2001. Cas1p is a membrane protein necessary for the O-acetylation of the Cryptococcus neoformans capsular polysaccharide. Mol. Microbiol. 42, 453–467. doi:10.1046/j.1365-2958.2001.02651.x

Jarvis, M.C., 2013. Cellulose Biosynthesis: Counting the Chains. PLANT Physiol. 163, 1485–1486. doi:10.1104/pp.113.231092

Jensen, J.K., Johnson, N.R., Wilkerson, C.G., 2014. Arabidopsis thaliana IRX10 and two related proteins from psyllium and Physcomitrella patens are xylan xylosyltransferases. Plant J. 80, 207–215. doi:10.1111/tpj.12641

Jensen, J.K., Kim, H., Cocuron, J.C., Orler, R., Ralph, J., Wilkerson, C.G., 2011. The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan

synthesis in Arabidopsis. Plant J. 66, 387–400. doi:10.1111/j.1365-313X.2010.04475.x

Jiang, N., Wiemels, R.E., Soya, A., Whitley, R., Held, M., Faik, A., 2016. Composition, Assembly, and Trafficking of a Wheat Xylan Synthase Complex. Plant Physiol. 170, 1999–2023. doi:10.1104/pp.15.01777

Kabel, M.A., De Waard, P., Schols, H.A., Voragen, A.G.J., 2003. Location of O-acetyl substituents in xylo-oligosaccharides obtained from hydrothermally treated Eucalyptus wood. Carbohydr. Res. 338, 69–77. doi:10.1016/S0008-6215(02)00351-8

Kauss, H., Hassid, W.Z., 1967. Biosynthesis of the 4-O-methyl-D-glucuronic acid unit of hemicellulose B by transmethylation from S-adenosyl-L-methionine. J. Biol. Chem. 242, 1680–1684.

Kellett, L.E., Poole, D.M., Ferreira, L.M., Durrant, A.J., Hazlewood, G.P., Gilbert, H.J., 1990. Xylanase B and an arabinofuranosidase from Pseudomonas fluorescens subsp. cellulosa contain identical cellulose-binding domains and are encoded by adjacent genes. Biochem. J. 272, 369–376. doi:10.1042/BJ2720369

Kennedy, J.F., Sandhu, J.S., Southgate, D.A.T., 1979. Structural data for the carbohydrate of ispaghula husk ex Plantago ovata forsk. Carbohydr. Res. 75, 265–274.

Keppler, B.D., Showalter, A.M., 2010. IRX14 and IRX14-LIKE, two glycosyl transferases involved in glucuronoxylan biosynthesis and drought tolerance in arabidopsis. Mol. Plant 3, 834–841. doi:10.1093/mp/ssq028

Kiefer, L.L., York, W.S., Darvill, A.G., Albersheim, P., 1989. Xyloglucan isolated from suspension-cultured sycamore cell walls is O-acetylated. Phytochemistry 28, 2105–2107. doi:10.1016/S0031-9422(00)97928-7

Kieliszewski, M.J., 2001. The latest hype on Hyp-O-glycosylation codes. Phytochemistry 57, 319–323. doi:10.1016/S0031-9422(01)00029-2

Kimura, S., 1999. Immunogold Labeling of Rosette Terminal Cellulose-Synthesizing Complexes in the Vascular Plant Vigna angularis. Plant Cell 11, 2075–2086. doi:10.1105/tpc.11.11.2075

Klepikova, A. V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., Penin, A.A., 2016. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. Plant J. 88, 1058–1070. doi:10.1111/tpj.13312

Komalavilas, P., Mort, A.J., 1989. The acetylation of O-3 of galacturonic acid in the rhamnose-rich portion of pectins. Carbohydr. Res. 189, 261–272. doi:10.1016/0008-6215(89)84102-3

Kotake, T., Tsuchiya, K., Aohara, T., Konishi, T., Kaneko, S., Igarashi, K., Samejima, M., Tsumuraya, Y., 2006. An α-L-arabinofuranosidase/β-D-xylosidase from immature seeds of radish (Raphanus sativus L.). J. Exp. Bot. 57, 2353–2362. doi:10.1093/jxb/erj206

Laubinger, S., Zeller, G., Henz, S.R., Sachsenberg, T., Widmer, C.K., Naouar, N., Vuylsteke, M., Schölkopf, B., Rätsch, G., Weigel, D., 2008. At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. Genome Biol. 9, R112. doi:10.1186/gb-2008-9-7-r112

Lavanant, H., Loutelier-Bourhis, C., 2012. Use of procaine and procainamide as derivatizing co-matrices for the analysis of oligosaccharides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 26, 1311–1319. doi:10.1002/rcm.6223

Le Goff, A., Renard, C.M.G.C., Bonnin, E., Thibault, J.F., 2001. Extraction, purification and chemical characterisation of xylogalacturonans from pea hulls. Carbohydr. Polym. 45, 325–334. doi:10.1016/S0144-8617(00)00271-X

Lee, C., Teng, Q., Huang, W., Zhong, R., Ye, Z.-H., 2010. The Arabidopsis Family GT43 Glycosyltransferases Form Two Functionally Nonredundant Groups Essential for the Elongation of Glucuronoxylan Backbone. Plant Physiol. 153, 526–541. doi:10.1104/pp.110.155309

Lee, C., Teng, Q., Zhong, R., Ye, Z.-H., 2012. Arabidopsis GUX proteins are glucuronyltransferases responsible for the addition of glucuronic acid side chains onto xylan. Plant Cell Physiol. 53, 1204–1216. doi:10.1093/pcp/pcs064

Lee, K.J.D., Marcus, S.E., Knox, J.P., 2011. Cell wall

biology: Perspectives from cell wall imaging. Mol. Plant 4, 212–219. doi:10.1093/mp/ssq075

Lee, H. V., Hamid, S.B.A., Zain, S.K., 2014. Conversion of lignocellulosic biomass to nanocellulose: Structure and chemical process. Sci. World J. 2014, 1–20. doi:10.1155/2014/631013

Lefebvre, V., Fortabat, M.-N., Ducamp, A., North, H.M., Maia-Grondard, A., Trouverie, J., Boursiac, Y., Mouille, G., Durand-Tardif, M., 2011. ESKIMO1 disruption in Arabidopsis alters vascular tissue and impairs water transport. PLoS One 6, e16645. doi:10.1371/journal.pone.0016645

Li, S., Bashline, L., Lei, L., Gu, Y., 2014. Cellulose Synthesis and Its Regulation. Arab. B. 12, e0169. doi:10.1199/tab.0169

Li, X., 2011. Development and Application of A New Method for Analysing Plant Cell Walls. University of Cambridge.

Li, X., Jackson, P., Rubtsov, D. V, Faria-Blanc, N., Mortimer, J.C., Turner, S.R., Krogh, K.B., Johansen, K.S., Dupree, P., 2013. Development and application of a high throughput carbohydrate profiling technique for analyzing plant cell wall polysaccharides and carbohydrate active enzymes. Biotechnol. Biofuels 6, 94. doi:10.1186/1754-6834-6-94

Liners, F., Thibault, J.F., Van Cutsem, P., 1992. Influence of the degree of polymerization of oligogalacturonates and of esterification pattern of pectin on their recognition by monoclonal antibodies. Plant Physiol. 99, 1099–104. doi:10.1104/PP.99.3.1099

Loix, C., Huybrechts, M., Vangronsveld, J., Gielen, M., Keunen, E., Cuypers, A., 2017. Reciprocal Interactions between Cadmium-Induced Cell Wall Responses and Oxidative Stress in Plants. Front. Plant Sci. 8, 1867. doi:10.3389/fpls.2017.01867

Lugan, R., Niogret, M.F., Kervazo, L., Larher, F.R., Kopka, J., Bouchereau, A., 2009. Metabolome and water status phenotyping of Arabidopsis under abiotic stress cues reveals new insight into ESK1 function. Plant, Cell Environ. 32, 95–108. doi:10.1111/j.1365-3040.2008.01898.x

Łyczakowski, J.J., 2018. Biosynthesis and function of glucuronic acid substitution patterns on softwood xylan. University of Cambridge.

Lyczakowski, J.J., Wicher, K.B., Terrett, O.M., Faria-Blanc, N., Yu, X., Brown, D., Krogh, K.B.R.M., Dupree, P., Busse-Wicher, M., 2017. Removal of glucuronic acid from xylan is a strategy to improve the conversion of plant biomass to sugars for bioenergy. Biotechnol. Biofuels 10, 224. doi:10.1186/s13068-017-0902-1

Manabe, Y., Nafisi, M., Verhertbruggen, Y., Orfila, C., Gille, S., Rautengarten, C., Cherk, C., Marcus, S.E., Somerville, S., Pauly, M., Knox, J.P., Sakuragi, Y., Scheller, H.V., 2011. Loss-of-Function Mutation of REDUCED WALL ACETYLATION2 in Arabidopsis Leads to Reduced Cell Wall Acetylation and Increased Resistance to Botrytis cinerea. Plant Physiol. 155, 1068–1078. doi:10.1104/pp.110.168989

Manabe, Y., Verhertbruggen, Y., Gille, S., Harholt, J., Chong, S.-L., Pawar, P.M.-A., Mellerowicz, E.J., Tenkanen, M., Cheng, K., Pauly, M., Scheller, H.V., 2013. Reduced Wall Acetylation Proteins Play Vital and Distinct Roles in Cell Wall O-Acetylation in Arabidopsis. Plant Physiol. 163, 1107–1117. doi:10.1104/pp.113.225193

Manfield, I.W., Orfila, C., McCartney, L., Harholt, J., Bernal, A.J., Scheller, H.V., Gilmartin, P.M., Mikkelsen, J.D., Knox, J.P., Willats, W.G.T., 2004. Novel cell wall architecture of isoxaben-habituated Arabidopsis suspension-cultured cells: Global transcript profiling and cellular analysis. Plant J. 40, 260–275. doi:10.1111/j.1365-313X.2004.02208.x

Manna, S., McAnalley, B.H., 1993. Determination of the position of the O-acetyl group in a β-(1 → 4)-mannan (acemannan) from Aloe barbardensis Miller. Carbohydr. Res. 241, 317–319. doi:10.1016/0008-6215(93)80122-U

Martínez-Abad, A., Berglund, J., Toriz, G., Gatenholm, P., Henriksson, G., Lindström, M., Wohlert, J., Vilaplana, F., 2017. Regular Motifs in Xylan Modulate Molecular Flexibility and Interactions with Cellulose Surfaces. Plant Physiol. 175, 1579–1592. doi:10.1104/pp.17.01184

Maslen, S.L., Goubet, F., Adam, A., Dupree, P., Stephens, E., 2007. Structure elucidation of arabinoxylan isomers by normal phase HPLC-MALDI-TOF/TOF-

MS/MS. Carbohydr. Res. 342, 724–735. doi:10.1016/j.carres.2006.12.007

Mastihubová, M., Biely, P., 2004. Lipase-catalysed preparation of acetates of 4-nitrophenyl β-D-xylopyranoside and their use in kinetic studies of acetyl migration. Carbohydr. Res. 339, 1353–1360. doi:10.1016/j.carres.2004.02.016

Matsunaga, T., Ishii, T., Matsumoto, S., Higuchi, M., Darvill, A.G., Albersheim, P., O'Neill, M.A., 2004. Occurrence of the Primary Cell Wall Polysaccharide Rhamnogalacturonan II in Pteridophytes, Lycophytes, and Bryophytes. Implications for the Evolution of Vascular Plants. Plant Physiol. 134, 339–351. doi:10.1104/pp.103.030072

Mazeau, K., Heux, L., 2003. Molecular dynamics simulations of bulk native crystalline and amorphous structures of cellulose. J. Phys. Chem. B 107, 2394–2403. doi:10.1021/jp0219395

Mazeau, K., Moine, C., Krausz, P., Gloaguen, V., 2005. Conformational analysis of xylan chains. Carbohydr. Res. 340, 2752–2760. doi:10.1016/j.carres.2005.09.023

McCann, M.C., Carpita, N.C., 2015. Biomass recalcitrance: A multi-scale, multi-factor, and conversion-specific property. J. Exp. Bot. doi:10.1093/jxb/erv267

Mohnen, D., 2008. Pectin structure and biosynthesis. Curr. Opin. Plant Biol. doi:10.1016/j.pbi.2008.03.006

Mølgaard, A., Larsen, S., 2004. Crystal packing in two pH-dependent crystal forms of rhamnogalacturonan acetylesterase. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 472–478. doi:10.1107/S0907444903029767

Mortimer, J.C., Faria-Blanc, N., Yu, X., Tryfona, T., Sorieul, M., Ng, Y.Z., Zhang, Z., Stott, K., Anders, N., Dupree, P., 2015. An unusual xylan in Arabidopsis primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. Plant J. 83, 413–426. doi:10.1111/tpj.12898

Mortimer, J.C., Miles, G.P., Brown, D.M., Zhang, Z., Segura, M.P., Weimar, T., Yu, X., Seffen, K.A., Stephens, E., Turner, S.R., Dupree, P., 2010. Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. Proc. Natl. Acad. Sci. 107, 17409–17414. doi:10.1073/pnas.1005456107

Mottiar, Y., Vanholme, R., Boerjan, W., Ralph, J., Mansfield, S.D., 2016. Designer lignins: Harnessing the plasticity of lignification. Curr. Opin. Biotechnol. doi:10.1016/j.copbio.2015.10.009

Nakamura, A., Furuta, H., Maeda, H., Takao, T., Nagamatsu, Y., 2002. Analysis of the molecular construction of xylogalacturonan isolated from soluble soybean polysaccharides. Biosci. Biotechnol. Biochem. 66, 1155–8. doi:10.1271/bbb.66.1155

Nelson, D., Cox, M., 2013. Carbohydrates and Glycobiology., in: Winslow, S. (Ed.), Lehninger Principles of Biochemistry. W.H. Freeman and Company, New York, pp. 243–280.

Neue, U.D., Phoebe, C.H., 1997. High-Performance Liquid Chromatography of Derivatized and Non-Derivatized Oligosaccharides: A Review, in: A Laboratory Guide to Glycoconjugate Analysis. Birkhäuser Basel, Basel, pp. 1–22. doi:10.1007/978-3-0348-7388-8_1

Newman, R.H., Hill, S.J., Harris, P.J., 2013. Wide-Angle X-Ray Scattering and Solid-State Nuclear Magnetic Resonance Data Combined to Test Models for Cellulose Microfibrils in Mung Bean Cell Walls. PLANT Physiol. 163, 1558–1567. doi:10.1104/pp.113.228262

Nikolovski, N., Rubtsov, D., Segura, M.P., Miles, G.P., Stevens, T.J., Dunkley, T.P.J., Munro, S., Lilley, K.S., Dupree, P., 2012. Putative Glycosyltransferases and Other Plant Golgi Apparatus Proteins Are Revealed by LOPIT Proteomics. PLANT Physiol. 160, 1037–1051. doi:10.1104/pp.112.204263

Nishiyama, Y., Langan, P., Chanzy, H., 2002. Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. J. Am. Chem. Soc. 124, 9074–9082. doi:10.1021/ja0257319

Nishiyama, Y., Sugiyama, J., Chanzy, H., Langan, P., 2003. Crystal Structure and Hydrogen Bonding System in Cellulose Iα from Synchrotron X-ray and Neutron Fiber Diffraction. J. Am. Chem. Soc. 125, 14300–14306. doi:10.1021/ja037055w

Nixon, B.T., Mansouri, K., Singh, A., Du, J., Davis, J.K.,

Lee, J.G., Slabaugh, E., Vandavasi, V.G., O'Neill, H., Roberts, E.M., Roberts, A.W., Yingling, Y.G., Haigler, C.H., 2016. Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex. Sci. Rep. 6. doi:10.1038/srep28696

O'Neill, M., Albersheim, P., Darvill, A.G., 1990. The Pectic Polysaccharides of Primary Cell Walls, in: Methods in Plant Biochemistry. Academic Press, pp. 415–441. doi:10.1016/B978-0-12-461012-5.50018-5

O'Neill, M.A., Ishii, T., Albersheim, P., Darvill, A.G., 2004. RHAMNOGALACTURONAN II: Structure and Function of a Borate Cross-Linked Cell Wall Pectic Polysaccharide. Annu. Rev. Plant Biol. 55, 109–139. doi:10.1146/annurev.arplant.55.031903.141750

Oehme, D.P., Downton, M.T., Doblin, M.S., Wagner, J., Gidley, M.J., Bacic, A., 2015. Unique Aspects of the Structure and Dynamics of Elementary Iβ Cellulose Microfibrils Revealed by Computational Simulations. Plant Physiol. 168, 3–17. doi:10.1104/pp.114.254664

Panchy, N., Lehti-Shiu, M.D., Shiu, S.-H., 2016. Evolution of gene duplication in plants. Plant Physiol. 171, 2294–2316. doi:10.1104/pp.16.00523

Patron, N.J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C., Smith, A., Hibberd, J., Webb, A.A.R., Locke, J., Schornack, S., Ajioka, J., Baulcombe, D.C., Zipfel, C., Kamoun, S., Jones, J.D.G., Kuhn, H., Robatzek, S., Van Esse, H.P., Sanders, D., Oldroyd, G., Martin, C., Field, R., O'Connor, S., Fox, S., Wulff, B., Miller, B., Breakspear, A., Radhakrishnan, G., Delaux, P.M., Loqué, D., Granell, A., Tissier, A., Shih, P., Brutnell, T.P., Quick, W.P., Rischer, H., Fraser, P.D., Aharoni, A., Raines, C., South, P.F., Ané, J.M., Hamberger, B.R., Langdale, J., Stougaard, J., Bouwmeester, H., Udvardi, M., Murray, J.A.H., Ntoukakis, V., Schäfer, P., Denby, K., Edwards, K.J., Osbourn, A., Haseloff, J., 2015. Standards for plant synthetic biology: A common syntax for exchange of DNA parts. New Phytol. 208, 13–19. doi:10.1111/nph.13532

Pauly, M., 1999. Development of analytical tools to study plant cell wall xyloglucan. Shaker.

Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A., Xiong, G., 2013. Hemicellulose biosynthesis. Planta. doi:10.1007/s00425-013-1921-1

Pauly, M., Keegstra, K., 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J. doi:10.1111/j.1365-313X.2008.03463.x

Pawar, P.M.-A., Koutaniemi, S., Tenkanen, M., Mellerowicz, E.J., 2013. Acetylation of woody lignocellulose: significance and regulation. Front. Plant Sci. 4, 118. doi:10.3389/fpls.2013.00118

Peña, M.J., Kulkarni, A.R., Backe, J., Boyd, M., O'Neill, M.A., York, W.S., 2016. Structural diversity of xylans in the cell walls of monocots. Planta 244, 589–606. doi:10.1007/s00425-016-2527-1

Peña, M.J., Zhong, R., Zhou, G.-K., Richardson, E.A., O'Neill, M.A., Darvill, A.G., York, W.S., Ye, Z.-H., 2007. Arabidopsis irregular xylem8 and irregular xylem9: Implications for the Complexity of Glucuronoxylan Biosynthesis. Plant Cell Online 19, 549–563. doi:10.1105/tpc.106.049320

Persson, S., Paredez, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., Khitrov, N., Auer, M., Somerville, C.R., 2007. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. Proc. Natl. Acad. Sci. 104, 15566–15571. doi:10.1073/pnas.0706592104

Popper, Z.A., Fry, S.C., 2008. Xyloglucan-pectin linkages are formed intra-protoplasmically, contribute to wall-assembly, and remain stable in the cell wall. Planta 227, 781–794. doi:10.1007/s00425-007-0656-2

Potikha, T., Delmer, D.P., 1995. A mutant of Arabidopsis thaliana displaying altered patterns of cellulose deposition. Plant J. 7, 453–460. doi:10.1046/j.1365-313X.1995.7030453.x

Puchart, V., Biely, P., 2015. Redistribution of acetyl groups on the non-reducing end xylopyranosyl residues and their removal by xylan deacetylases. Appl. Microbiol. Biotechnol. 99, 3865–3873. doi:10.1007/s00253-014-6160-2

Ralet, M.-C., Crépeau, M.-J., Vigouroux, J., Tran, J., Berger, A., Sallé, C., Granier, F., Botran, L., North, H.M., 2016. Xylans Provide the Structural Driving Force for Mucilage Adhesion to the Arabidopsis Seed Coat. Plant Physiol. 171, 165–178.

doi:10.1104/pp.16.00211

Ramírez, V., Xiong, G., Mashiguchi, K., Yamaguchi, S., Pauly, M., 2018. Growth- and stress-related defects associated with wall hypoacetylation are strigolactone-dependent. Plant Direct 2, e00062. doi:10.1002/pld3.62

Ratke, C., Pawar, P.M.-A., Balasubramanian, V.K., Naumann, M., Duncranz, M.L., Derba-Maceluch, M., Gorzsás, A., Endo, S., Ezcurra, I., Mellerowicz, E.J., 2015. Populus GT43 family members group into distinct sets required for primary and secondary wall xylan biosynthesis and include useful promoters for wood modification. Plant Biotechnol. J. 13, 26–37. doi:10.1111/pbi.12232

Ratke, C., Terebieniec, B.K., Winestrand, S., Derba-Maceluch, M., Grahn, T., Schiffthaler, B., Ulvcrona, T., Özparpucu, M., Rüggeberg, M., Lundqvist, S.O., Street, N.R., Jönsson, L.J., Mellerowicz, E.J., 2018. Downregulating aspen xylan biosynthetic GT43 genes in developing wood stimulates growth via reprograming of the transcriptome. New Phytol. 219, 230–245. doi:10.1111/nph.15160

Reiter, W.D., Vanzin, G.F., 2001. Molecular genetics of nucleotide sugar interconversion pathways in plants. Plant Mol. Biol. doi:10.1023/A:1010671129803

Ren, Y., Hansen, S.F., Ebert, B., Lau, J., Scheller, H.V., 2014. Site-directed mutagenesis of IRX9, IRX9L and IRX14 proteins involved in xylan biosynthesis: Glycosyltransferase activity is not required for IRX9 function in arabidopsis. PLoS One 9, e105014. doi:10.1371/journal.pone.0105014

Rennie, E.A., Hansen, S.F., Baidoo, E.E.K., Hadi, M.Z., Keasling, J.D., Scheller, H.V., 2012. Three Members of the Arabidopsis Glycosyltransferase Family 8 Are Xylan Glucuronosyltransferases. Plant Physiol. 159, 1408–1417. doi:10.1104/pp.112.200964

Rennie, E.A., Scheller, H.V., 2014. Xylan biosynthesis. Curr. Opin. Biotechnol. doi:10.1016/j.copbio.2013.11.013

Ridlova, G., Mortimer, J.C., Maslen, S.L., Dupree, P., Stephens, E., 2008. Oligosaccharide relative quantitation using isotope tagging and normal-phase liquid chromatography/mass spectrometry. Rapid Commun. Mass Spectrom. 22, 2723–2730. doi:10.1002/rcm.3665

Ros-Barceló, A., Gómez-Ros, L. V, Ferrer, M.A., Hernández, J.A., 2006. The apoplastic antioxidant enzymatic system in the wood-forming tissues of trees. Trees - Struct. Funct. 20, 145–156. doi:10.1007/s00468-005-0020-8

Rubinstein, C. V., Gerrienne, P., de la Puente, G.S., Astini, R.A., Steemans, P., 2010. Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New Phytol. 188, 365–369. doi:10.1111/j.1469-8137.2010.03433.x

Sambrook, J., Russell, D.W., 2006. The Inoue Method for Preparation and Transformation of Competent E. Coli : “Ultra-Competent” Cells. Cold Spring Harb. Protoc. 2006, pdb.prot3944. doi:10.1101/pdb.prot3944

Samuelsen, A.B., Hanne Cohen, E., Smestad Paulsen, B., Brüll, L.P., Thomas-Oates, J.E., 1999. Structural studies of a heteroxylan from Plantago major L. seeds by partial hydrolysis, HPAEC-PAD, methylation and GC-MS, ESMS and ESMS/MS. Carbohydr. Res. 315, 312–318. doi:10.1016/S0008-6215(99)00038-5

Scheller, H.V., Jensen, J.K., Sørensen, S.O., Harholt, J., Geshi, N., 2007. Biosynthesis of pectin. Physiol. Plant. doi:10.1111/j.1399-3054.2006.00834.x

Scheller, H.V., Ulvskov, P., 2010. Hemicelluloses. Annu. Rev. Plant Biol. 61, 263–289. doi:10.1146/annurev-arplant-042809-112315

Schols, H.A., Bakx, E.J., Schipper, D., Voragen, A.G.J., 1995. A xylogalacturonan subunit present in the modified hairy regions of apple pectin. Carbohydr. Res. 279, 265–279. doi:10.1016/0008-6215(95)00287-1

Schultink, A., Naylor, D., Dama, M., Pauly, M., 2015. The Role of the Plant-Specific ALTERED XYLOGLUCAN9 Protein in Arabidopsis Cell Wall Polysaccharide *O*- Acetylation. Plant Physiol. 167, 1271–1283. doi:10.1104/pp.114.256479

Shimada, T.L., Shimada, T., Hara-Nishimura, I., 2010. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. Plant J. 61, 519–528. doi:10.1111/j.1365-313X.2009.04060.x

Showalter, A.M., 2001. Arabinogalactan-proteins: structure,

expression and function. Cell. Mol. Life Sci. 58, 1399–1417. doi:1420-682X/01/101399-19

Simmons, T.J., Mortimer, J.C., Bernardinelli, O.D., Pöppler, A.-C., Brown, S.P., deAzevedo, E.R., Dupree, R., Dupree, P., 2016. Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. Nat. Commun. 7, 13902. doi:10.1038/ncomms13902

Somerville, C., 2006. Cellulose Synthesis in Higher Plants. Annu. Rev. Cell Dev. Biol. 22, 53–78. doi:10.1146/annurev.cellbio.22.022206.160206

Song, L., Zeng, W., Wu, A., Picard, K., Lampugnani, E.R., Cheetamun, R., Beahan, C., Cassin, A., Lonsdale, A., Doblin, M.S., Bacic, A., 2015. Asparagus spears as a model to study heteroxylan biosynthesis during secondary wall development. PLoS One 10, e0123878. doi:10.1371/journal.pone.0123878

Sørensen, I., Pettolino, F.A., Wilson, S.M., Doblin, M.S., Johansen, B., Bacic, A., Willats, W.G.T., 2008. Mixed-linkage (1 → 3),(1 → 4)-β-D-glucan is not unique to the Poales and is an abundant component of Equisetum arvense cell walls. Plant J. 54, 510–521. doi:10.1111/j.1365-313X.2008.03453.x

St John, F.J., Hurlbert, J.C., Rice, J.D., Preston, J.F., Pozharski, E., 2011. Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. J. Mol. Biol. 407, 92–109. doi:10.1016/j.jmb.2011.01.010

Stranne, M., Ren, Y., Fimognari, L., Birdseye, D., Yan, J., Bardor, M., Mollet, J.-C., Komatsu, T., Kikuchi, J., Scheller, H.V., Sakuragi, Y., 2018. *TBL10* is required for *O* -acetylation of pectic rhamnogalacturonan-I in *Arabidopsis thaliana*. Plant J. 0–2. doi:10.1111/tpj.14067

Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., Hao, Z., Zhu, X., Avci, U., Miller, J.S., Baldwin, D., Pham, C., Orlando, R., Darvill, A., Hahn, M.G., Kieliszewski, M.J., Mohnen, D., 2013. An *Arabidopsis* Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein. Plant Cell 25, 270–287. doi:10.1105/tpc.112.107334

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K., Turner, S.R., 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. 100, 1450–1455. doi:10.1073/pnas.0337628100

Teleman, A., Lundqvist, J., Tjerneld, F., Stålbrand, H., Dahlman, O., 2000. Characterization of acetylated 4-O-methylglucuronoxylan isolated from aspen employing 1H and 13C NMR spectroscopy. Carbohydr. Res. 329, 807–815. doi:10.1016/S0008-6215(00)00249-4

Teleman, A., Tenkanen, M., Jacobs, A., Dahlman, O., 2002. Characterization of O-acetyl-(4-O-methylglucurono)xylan isolated from birch and beech. Carbohydr. Res. 337, 373–377. doi:10.1016/S0008-6215(01)00327-5

Tenkanen, M., Luonteri, E., Teleman, A., 1996. Effect of side groups on the action of β-xylosidase from Trichoderma reesei against substituted xylo-oligosaccharides. FEBS Lett. 399, 303–306. doi:10.1016/S0014-5793(96)01313-0

Terrett, O.M., Dupree, P., 2019. Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. Curr. Opin. Biotechnol. doi:10.1016/j.copbio.2018.10.010

Tryfona, T., Liang, H.-C., Kotake, T., Tsumuraya, Y., Stephens, E., Dupree, P., 2012. Structural Characterization of Arabidopsis Leaf Arabinogalactan Polysaccharides. Plant Physiol. 160, 653–666. doi:10.1104/pp.112.202309

Tryfona, T., Liang, H.C., Kotake, T., Kaneko, S., Marsh, J., Ichinose, H., Lovegrove, A., Tsumuraya, Y., Shewry, P.R., Stephens, E., Dupree, P., 2010. Carbohydrate structural analysis of wheat flour arabinogalactan protein. Carbohydr. Res. 345, 2648–2656. doi:10.1016/j.carres.2010.09.018

Urbániková, Ľ., Vršanská, M., Mørkeberg Krogh, K.B.R., Hoff, T., Biely, P., 2011. Structural basis for substrate recognition by Erwinia chrysanthemi GH30 glucuronoxylanase. FEBS J. 278, 2105–2116. doi:10.1111/j.1742-4658.2011.08127.x

Urbanowicz, B.R., Peña, M.J., Moniz, H.A., Moremen, K.W., York, W.S., 2014. Two Arabidopsis proteins synthesize acetylated xylan in vitro. Plant J. 80, 197–206. doi:10.1111/tpj.12643

Urbanowicz, B.R., Peña, M.J., Ratnaparkhe, S., Avci, U., Backe, J., Steet, H.F., Foston, M., Li, H., O'Neill, M.A., Ragauskas, A.J., Darvill, A.G., Wyman, C.,

Gilbert, H.J., York, W.S., 2012. 4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. Proc. Natl. Acad. Sci. 109, 14253–14258. doi:10.1073/pnas.1208097109

Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., Vandepoele, K., 2012. Dissecting Plant Genomes with the PLAZA Comparative Genomics Platform. PLANT Physiol. 158, 590–600. doi:10.1104/pp.111.189514

Vardakou, M., Dumon, C., Murray, J.W., Christakopoulos, P., Weiner, D.P., Juge, N., Lewis, R.J., Gilbert, H.J., Flint, J.E., 2008. Understanding the Structural Basis for Substrate and Inhibitor Recognition in Eukaryotic GH11 Xylanases. J. Mol. Biol. 375, 1293–1305. doi:10.1016/j.jmb.2007.11.007

Vega-Sanchez, M.E., Verhertbruggen, Y., Christensen, U., Chen, X., Sharma, V., Varanasi, P., Jobling, S.A., Talbot, M., White, R.G., Joo, M., Singh, S., Auer, M., Scheller, H. V., Ronald, P.C., 2012. Loss of Cellulose Synthase-Like F6 Function Affects Mixed-Linkage Glucan Deposition, Cell Wall Mechanical Properties, and Defense Responses in Vegetative Tissues of Rice. PLANT Physiol. 159, 56–69. doi:10.1104/pp.112.195495

Voiniciuc, C., Guenl, M., Schmidt, M.H.-W., Usadel, B., 2015. Highly Branched Xylan Made by IRX14 and MUCI21 Links Mucilage to Arabidopsis Seeds. Plant Physiol. 169, pp.01441.2015. doi:10.1104/pp.15.01441

Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., Cumming, M., Kelley, L., Sternberg, M., Krishnakumar, V., Ferlanti, E., Miller, J., Town, C., Stuerzlinger, W., Provart, N.J., 2017. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. Plant Cell tpc.00073.2017. doi:10.1105/tpc.17.00073

Watanabe, T., Koshijima, T., 1988. Evidence for an Ester Linkage between Lignin and Glucuronic Acid in Lignin-Carbohydrate Complexes by DDQ-Oxidation. Biol, Chem 52, 2953–2955. doi:10.1080/00021369.1988.10869116

Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S., 2011. A modular cloning system for standardized assembly of multigene constructs. PLoS One 6, 16765. doi:10.1371/journal.pone.0016765

Weigel, D., Glazebrook, J., 2006. Setting Up Arabidopsis Crosses. Cold Spring Harb. Protoc. 2006, pdb.prot4623-pdb.prot4623. doi:10.1101/pdb.prot4623

Wellman, C.H., Osterloff, P.L., Mohiuddin, U., 2003. Fragments of the earliest land plants. Nature 425, 282–285. doi:10.1038/nature01884

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., Provart, N.J., 2007. An 'electronic fluorescent pictograph' Browser for exploring and analyzing large-scale biological data sets. PLoS One 2, e718. doi:10.1371/journal.pone.0000718

Wu, A.-M., Hornblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P., Marchant, A., 2010. Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L Pairs of Glycosyltransferase Genes Reveals Critical Contributions to Biosynthesis of the Hemicellulose Glucuronoxylan. Plant Physiol. 153, 542–554. doi:10.1104/pp.110.154971

Wu, A.-M., Rihouey, C., Seveno, M., Hörnblad, E., Singh, S.K., Matsunaga, T., Ishii, T., Lerouge, P., Marchant, A., 2009. The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. Plant J. 57, 718–731. doi:10.1111/j.1365-313X.2008.03724.x

Xin, Z., Browse, J., 1998. eskimo1 mutants of Arabidopsis are constitutively freezing-tolerant. Proc. Natl. Acad. Sci. 95, 7799–7804. doi:10.1073/pnas.95.13.7799

Xin, Z., Mandaokar, A., Chen, J., Last, R.L., Browse, J., 2007. Arabidopsis ESK1 encodes a novel regulator of freezing tolerance. Plant J. 49, 786–799. doi:10.1111/j.1365-313X.2006.02994.x

Xiong, G., Cheng, K., Pauly, M., 2013. Xylan O-acetylation impacts xylem development and enzymatic recalcitrance as indicated by the arabidopsis mutant tbl29. Mol. Plant 6, 1373–1375. doi:10.1093/mp/sst014

York, W.S., Darvill, A.G., Albersheim, P., 1984. Inhibition of 2,4-dichlorophenoxyacetic Acid-stimulated elongation of pea stem segments by a xyloglucan oligosaccharide. Plant Physiol. 75, 295–297.

doi:10.1104/pp.75.2.295

Yuan, Y., Teng, Q., Lee, C., Zhong, R., Ye, Z.H., 2014. Modification of the degree of 4-O-methylation of secondary wall glucuronoxylan. Plant Sci. 219–220, 42–50. doi:10.1016/j.plantsci.2014.01.005

Yuan, Y., Teng, Q., Zhong, R., Haghighat, M., Richardson, E.A., Ye, Z.-H., 2016a. Mutations of Arabidopsis TBL32 and TBL33 affect xylan acetylation and secondary wall deposition. PLoS One 11, e0146460. doi:10.1371/journal.pone.0146460

Yuan, Y., Teng, Q., Zhong, R., Ye, Z.-H., 2016b. Roles of Arabidopsis TBL34 and TBL35 in xylan acetylation and plant growth. Plant Sci. 243, 120–130. doi:10.1016/j.plantsci.2015.12.007

Yuan, Y., Teng, Q., Zhong, R., Ye, Z.-H., 2013. The arabidopsis DUF231 domain-containing protein ESK1 mediates 2-O- and 3-O-acetylation of xylosyl residues in xylan. Plant Cell Physiol. 54, 1186–1199. doi:10.1093/pcp/pct070

Yuan, Y., Teng, Q., Zhong, R., Ye, Z.H., 2016c. TBL3 and TBL31, Two Arabidopsis DUF231 domain proteins, are required for 3-O-monoacetylation of xylan. Plant Cell Physiol. 57, 35–45. doi:10.1093/pcp/pcv172

Zablackis, E., Huang, J., Muller, B., Darvill, A.G., Albersheim, P., 1995. Characterization of the Cell-Wall Polysaccharides of Arabidopsis thaliana Leaves. Plant Physiol. 107, 1129–1138. doi:10.1104/pp.107.4.1129

Zeng, W., Lampugnani, E.R., Picard, K.L., Song, L., Wu, A.-M., Farion, I.M., Zhao, J., Ford, K., Doblin, M.S., Bacic, A., 2016. Asparagus IRX9, IRX10, and IRX14A Are Components of an Active Xylan Backbone Synthase Complex that Forms in the Golgi Apparatus. Plant Physiol. 171, 93–109. doi:10.1104/pp.15.01919

Zhang, B., Zhang, L., Li, F., Zhang, D., Liu, X., Wang, H., Xu, Z., Chu, C., Zhou, Y., 2017. Control of secondary cell wall patterning involves xylan deacetylation by a GDSL esterase. Nat. Plants, Publ. online 3 March 2017; | doi10.1038/nplants.2017.17 3, 17017. doi:10.1038/NPLANTS.2017.17

Zhong, R., Cui, D., Phillips, D.R., Ye, Z.H., 2018a. A Novel Rice Xylosyltransferase Catalyzes the Addition of 2-*O* -Xylosyl Side Chains onto the Xylan Backbone. Plant Cell Physiol. 59, 554–565. doi:10.1093/pcp/pcy003

Zhong, R., Cui, D., Ye, Z.-H., 2018b. A group of Populus trichocarpa DUF231 proteins exhibit differential Oacetyltransferase activities toward xylan. PLoS One 13. doi:10.1371/journal.pone.0194532

Zhong, R., Cui, D., Ye, Z.H., 2018c. Members of the DUF231 Family are O-Acetyltransferases Catalyzing 2-O- and 3-O-Acetylation of Mannan. Plant Cell Physiol. 59, 2339–2349. doi:10.1093/pcp/pcy159

Zhong, R., Cui, D., Ye, Z.H., 2017. Regiospecific acetylation of xylan is mediated by a group of DUF231-containing O-acetyltransferases. Plant Cell Physiol. 58, 2126–2138. doi:10.1093/pcp/pcx147

Zhu, X.F., Sun, Y., Zhang, B., Mansoori, N., Wan, J.X., Liu, Y., Wang, Z.W., Shi, Y.Z., Zhou, Y., Zheng, S.J., 2014. *TRICHOME BIREFRINGENCE-LIKE27* Affects Aluminum Sensitivity by Modulating the *O* -Acetylation of Xyloglucan and Aluminum-Binding Capacity in Arabidopsis. Plant Physiol. 166, 181–189. doi:10.1104/pp.114.243808

Annex

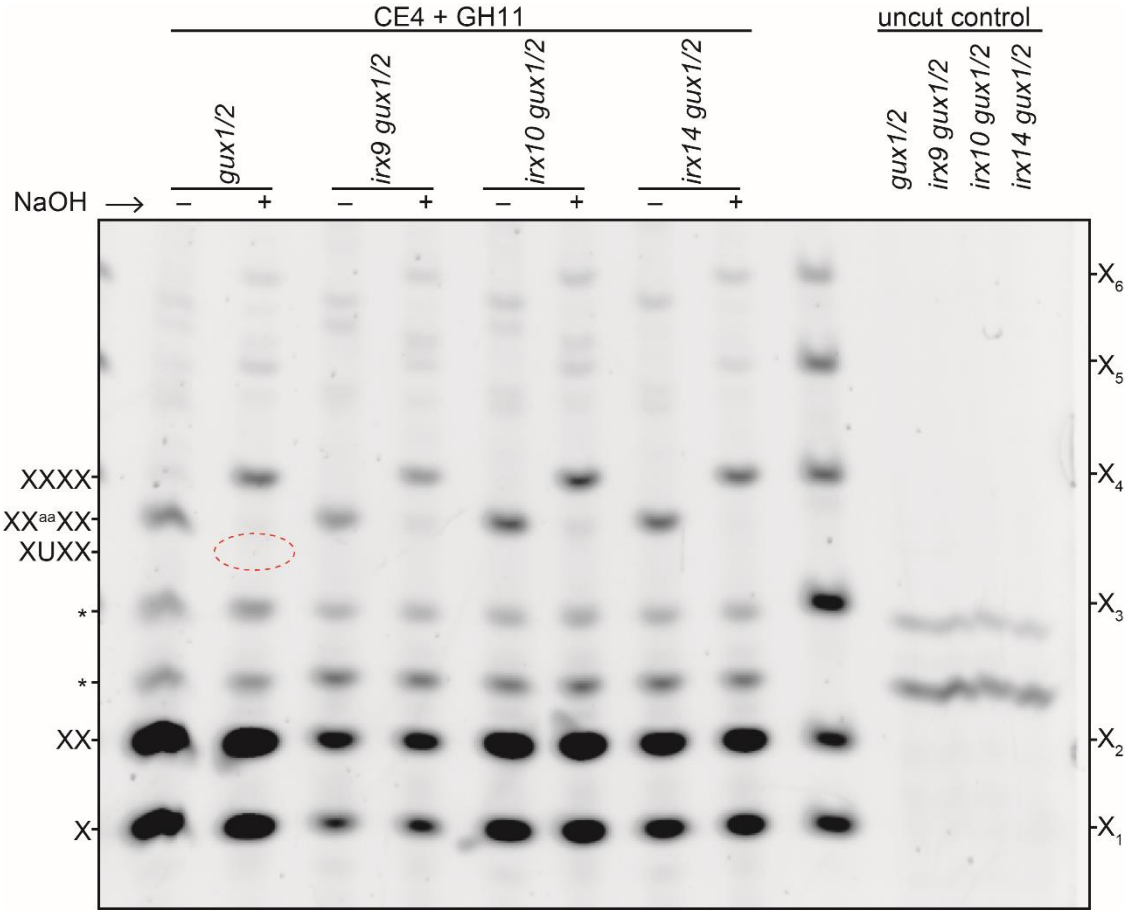


Figure S1: Analysis of [Me]GlcA on xylan in the *irx gux1/2* mutants.
Hollocellulose was coincubated with GH11 xylanase and CE4 acetylesterase. The enzymes were inactivated with heat and by purifying the samples with Nanosep. The oligosaccharides were then alkali-treated (+/-) and analysed on PACE. The dotted oval marks the position were the glucuronidated oligosaccharide XU$^{[me],a}$XX should migrate.

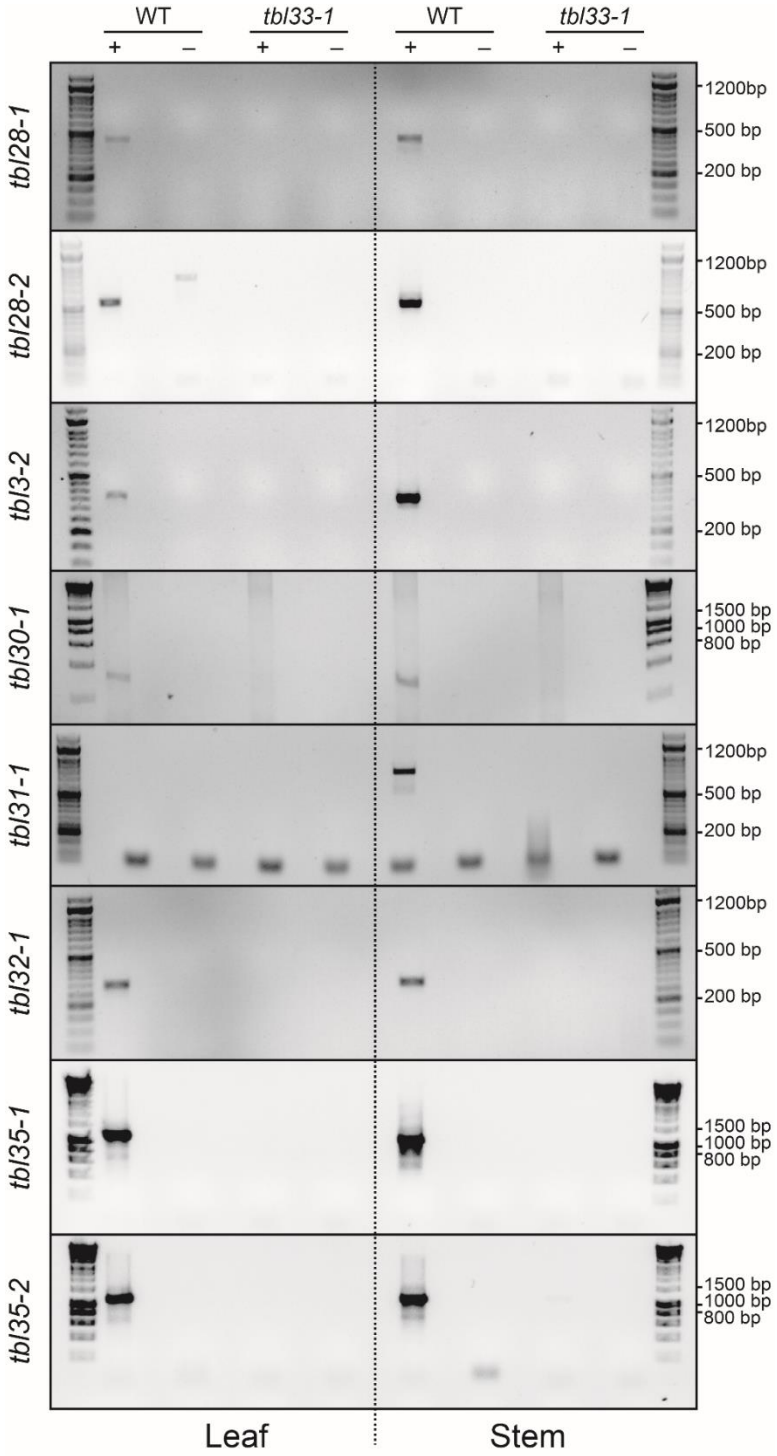
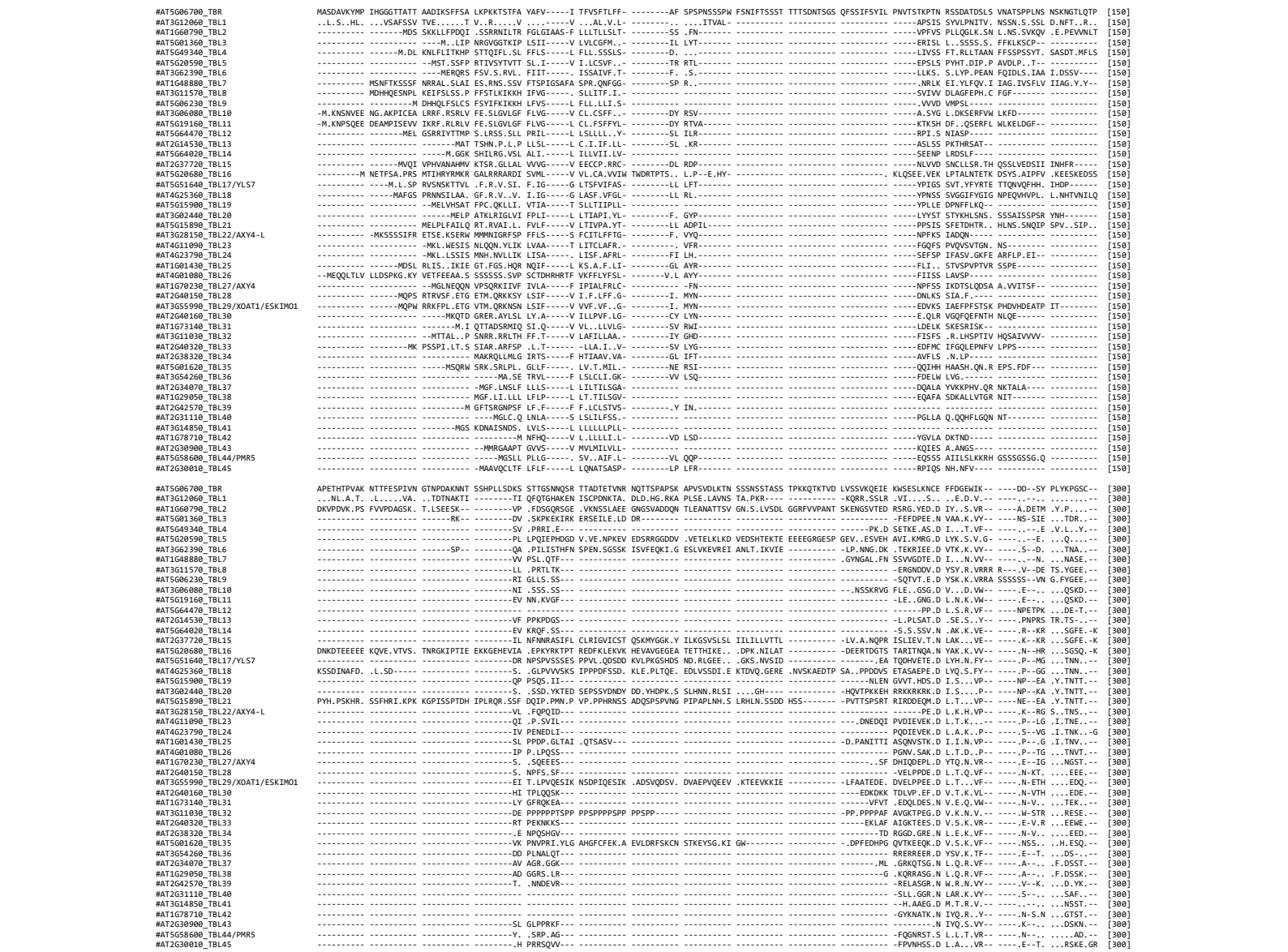


Figure S2: TBL genes expression analysis by RT-PCR.
Expression analysis of *TBL* genes in the *tbl* mutants by RT-PCR.
Leaf and stem tissues were harvested from 6-week-old plants, and RNA was extracted with a kit (Spectrum, Sigma-Aldrich), following the manufacturers' instructions. cDNA was produced from the template RNA using SuperScript™ III reverse transcriptase (Thermo fisher), reactions without retrotranscriptase controls (– RT) were used to identify genomic DNA contamination. cDNA samples were used for the PCR amplification of the respective *TBL* gene. For primer sequences see Table 2.6 in Chapter 2, page 65). The PCR products were resolved on a 1.25% agarose gel in TAE buffer. The DNA molecular weight markers used were 50bp DNA Ladder (RTU) and HyperLadder 1kb (Bioline).

Figure S3: Alignment of the *Arabidopsis* TBL protein family.
The 46 full amino acid sequences were aligned using the MUSCLE algorithm.

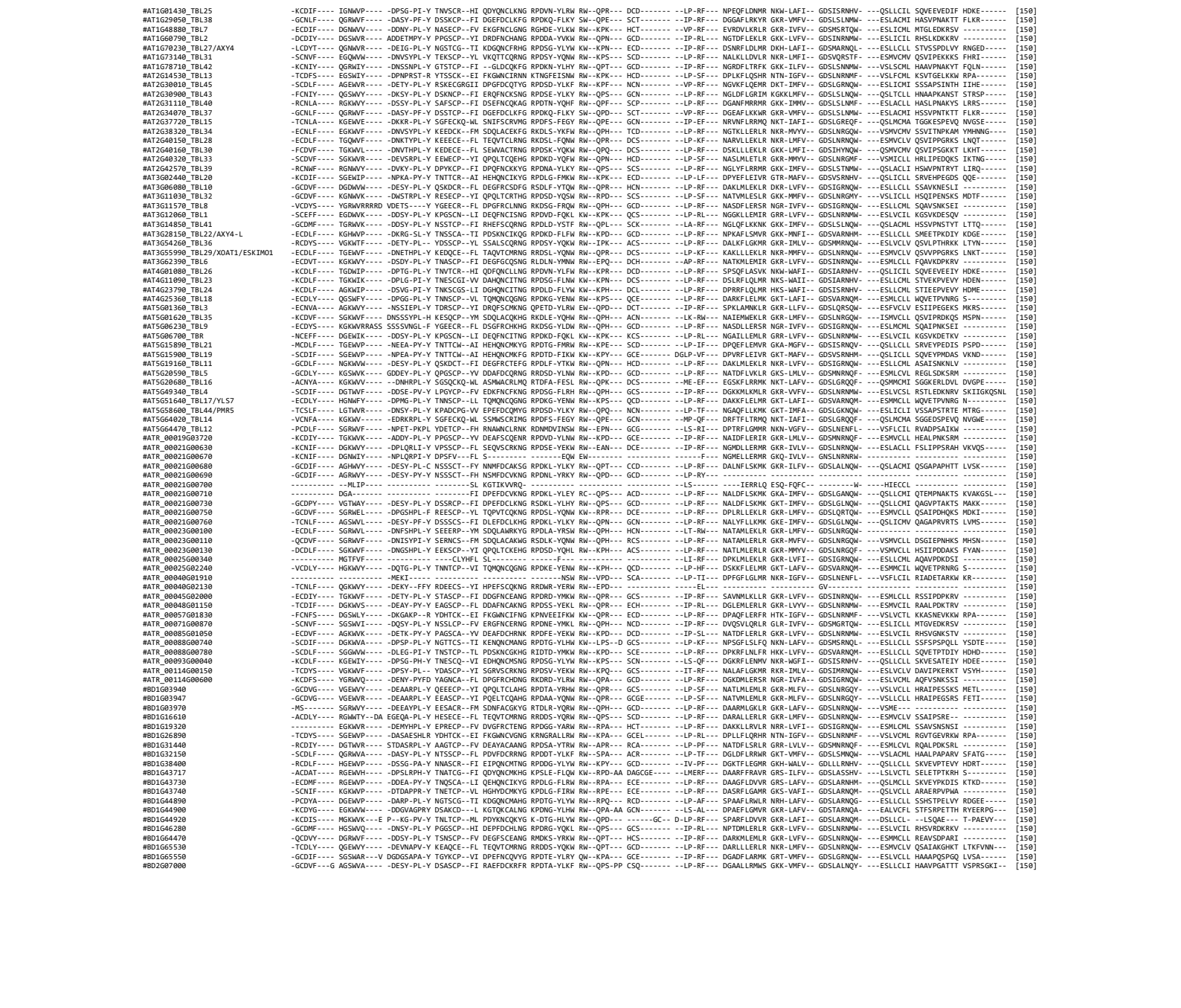


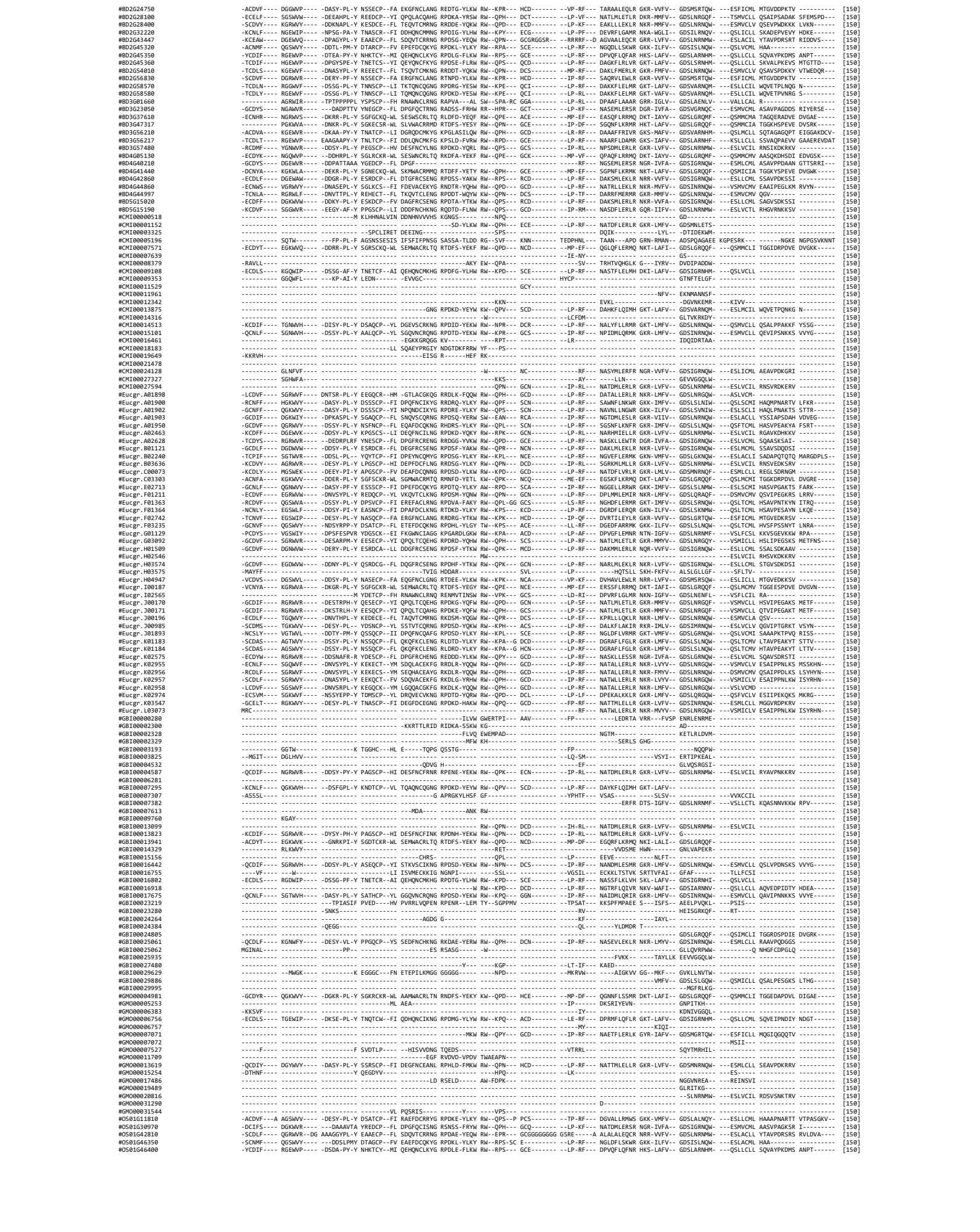
```
#AT5G06700_TBR                  GKS-SEQYKG A----DVIVF NTGH-WWTHE K-----TSKG E-DY-YQEGS N-VYHEL-AV LEAFRKALTT WGRWVEKNVN PAKSL-VFFR GYSASHFSGG QWNS------ ----GGACDS ETEPIKND-T YL-------T PYPSKM--KV  [600]
#AT3G12060_TBL1                 ....-..... .....IL.. ....-.---. .------.... ..-.-..... T-.HPK..-D. D..------- .....D.... .K.....-.. ...P..---- ...A------ ----....D ......E-. .-------. ..ML..--EI  [600]
#AT1G60790_TBL2                 D.T-TSM.RD .-----.ILI. .----.---D .-----.KL. .-N.----.N V--.PR..-K. ...YKR..I. .AK..D..ID RSQTH-IV. ...VT..R.. P...------ ----.-Q.HK .----.F.T-S .-------A K.....--.A  [600]
#AT5G01360_TBL3                 KDR-AKFWE. .-----.IL.. ..YV-..MSG LR----MKAL W-GS-FGN.E S-GAEA.-DT QV.Y.LG.K. .AN..DST.D .N.TR-...T TM.PT.TRSA D.GK------ -PN-.TK.FN ..K...DK-K FW------G TGSN.QMM..  [600]
#AT5G49340_TBL4                 QR.MTKI..N .-----.IVI. ....-....Q .-----.YE. K-G.-....N R--.ER.-E. K..YT..IH. .AD..DS.I. ST.TR-...V ...S...RK. A...------ ----.Q..G ..R..Q.E-. .T-------- -GVYPWMM..  [600]
#AT5G20590_TBL5                 D..-HAKW.R .-----.IL. .....-.V.G .-----.AR. K-N.-.K..D Y-I.PKF-DA T..Y.RS.K. .AK.IDQ... .K.Q.-..Y. ...SA..R.. E.D------ ----..S.NG .V..V.KG-S II------D S..L.----.I  [600]
#AT3G62390_TBL6                 DRT-.SRW.. .----NIL.. ..A.-..S.Y .-----.KS. V-N.-....D L-IHPK..-D. ST..K...Q. .SS..D...D .K.TR-.... SAAP..... E...------ ----..H.RE ANM.LNQT-- ---------F KPSYSSKKSI  [600]
#AT1G48880_TBL7                 DVI-NHEWSS .----.FLI. ...Q-..VPG .-----.LFET G-C-.F.V.N S-LRLGM-SI PA.Y.V..E. .AS.I.ST.D .N.TR-.L.. TFEP..---- -.SD------ ----HRS.NV TKY.APDT-- ---------E GRDKSIFSEM  [600]
#AT3G11570_TBL8                 NWQ-.KKWV. S----..L.. ....-.NED .-----.FIA G-C-.F...G K-LNKTM-G. M.G.E.S.K. .KS..LERLD SER.H-.... SF.PV.YRN. T..L------ ----..L..A D...ETDMKK ME------P DPIHNNYISQ  [600]
#AT5G06230_TBL9                 NWQ-.KRWV. S----..L.. .S...-.NED .-----.VLT G-C-.FE..R K-.NKTM-G. M...G.S.K. .KS..LEKLD .D..Y-.... S..PV.YRN. T..T------ ----..L..A .I..ETDKRK LE------P DASHNEYIYK  [600]
#AT3G06080_TBL10                DWT-.SKWRD .-----.LiL .....-.NEG .------.TRT G-C-.F...E F-.KLKM-N. DD.YKR..N. VVK.IHTELD SN.TQ-.... TFAPV..R.. D.KT------ ----..T.HM ..L.EIGT-S L-------A SSETWEQL.I  [600]
#AT5G19160_TBL11                EWT-ADKWRD .-----.IL.. .....-.NY. .------.IR. G-C-.F...E K-.RMRM-KI EH.Y.R.MK. VMK.IQEE.D AN.TQ-.... TFAPV..R.. D.RT------ ----..T.HM ..L.DFGASL VP------A ETWDHI--.L  [600]
#AT5G64470_TBL12                 ----ANEWIN VTSFY..LI. .S..-..GYD .F---PKET PLVF-.RK.K P-INPP-.DI .PG.ELV.QN MVSYIQRE.- ...T..K-.W. LQ.PR..Y.. D..Q------ ----N.S.LL DKPLEE.QLD LWFDPRNNGV NKEARKINQI  [600]
#AT2G14530_TBL13                 DIP-DSSWAK .SSFH.ILIL ....W..APS .F---DPV.S PMLF-FEG.R P-ILPPI-PP ATGLDRV.NN MVNF...TKR .GGI--I... TQ.PR..E.. D.DQ------ ----..T.Q- RLQ.LLPG-K VE-EFFSVGN NGTNVEVRL.  [600]
#AT5G64020_TBL14                 PAFMRNYLHR F----.H.L.L .----.H.NRG .I----EGNH W-VM-HVN.T QVEGEY.KDI RN.KDFTIHS VAK.LDAQLP LHPR.KA.. TI.PR..KN. D..T------ ----..N.N- N.V.LSRGSE IT-------G DDG.ID--AT  [600]
#AT2G37720_TBL15                 PAFIRNYLHR F----H.L.L .----.H.SRD .I----EKNH W-VM-HVN.T RVEGGYFKN. EN.KIFTIHS LVK.LDAQLP LHPR.KA.T TI.PR..---- ---------- -----EK.N- N.I.LSRGSK IT-------G EGG.LD--TI  [600]
#AT5G20680_TBL16                 PAFLRQYLQK I----..L.M ....-.H.NRG .L----NGNK W-VM-HVN.V PNTNRK.A.L GN.KNFTIHS TVS..NSQLP LHPG.KA.Y. SL.PR..V.. E..T------ ----..S.NN T.PMSIGKEV LQ------E ESSDYS----  [600]
#AT5G51640_TBL17/YLS7            DERII.ALPN F----..V.L SS...FAKQ SVYILNDQIV G-GQLWWPDK S-KPEKI-NN V...GISVE. IIKAMA.HP. YTGLT--IL. TW.PD.YE.. A..T------ ----..S.TG KV..LPPG-N LV-------. NGFTEIMHEK  [600]
#AT4G25360_TBL18                 DERIM.AIPK F----..V.L SS...FAKQ SVYILKEEIV G-GQLWWPDK S-KPMKV-NN VD..GISVE. ILKSMATHP. YSGLT--IV. TF.PD.YE.. A..T------ ----..S.TG KE...LPG-K LV-------K NGFTEIMHEK  [600]
#AT5G15900_TBL19                 DE.WTADIGD F----.FVII SS...HYRP SVYVENRTIT G-CH-.CQ-L P-NITD..TM FYGY...FR. AFKAILDSES FKGVM--YL. SFAP...E.. L..E------ ----..D.L- RKQ.YRSN-- ET-------Q DETTMKLH.I  [600]
#AT3G02440_TBL20                 DPTWAS.IGE F----.Y.II SS.Q-.FFRP LFLFDKQKRI G-CL-.CY-I P-GVRNV-GA HF.Y.R..R. TFKTILGLE. FKGE--..L. TFAP...E.. E.DK------ ---------- ---------- ---------G AMRQS*----  [600]
#AT5G15890_TBL21                 DTKWTS.LDQ L----.YL.I SS...FSRP VIFYENQQIS G-CQ-.CA-L P-NTT...PL TYGY....RI SLKAII--E. FKGLA--.L. SF.PQ..E.. A..E------ ----..D.V- R.Q.YRRN-- ET-------I .EADLKVHDI  [600]
#AT3G28150_TBL22/AXY4-L          DEGWFNGLPN T----.IAIV SAA.-.FFRP IFIHRGDETL G-CI-.CN-L P-NMTQI-SP E.G.KLVYSA VL.QINECEM CK.D.VTVL. TI.PA..EN. T.DT------ ----..T.S- R.S.FGEN-K ID------L QSNEMKIR.S  [600]
#AT4G11090_TBL23                 DNTWTDLFPS L-----.YAII SS.E-.FLKT AVYHENANPV G-CH-.GCPE S-NMTD..GF DY.YNTS.RH VMDFIA.S-- KT.GM-I... TSIPD..ED. E.HN------ ----..T.K- K...VGEE-A VE------M KVLN.ILRD.  [600]
#AT4G23790_TBL24                 DETWTSLMPS F----.YAII S..K-..FLKS AIYHENAKLV G-CH-NCQEK P-HIE...GF DY.YNAS.HN VMDFLAAED. SKGT--.... TSTPD..QN. E.H.------ ----..T.K- Q...VSDE-E IE------I KDVH.ILKDI  [600]
#AT1G01430_TBL25                 DQKWTD..IN F----.YV.I SG.K-.FLKT TIFHENNTVT G-CH-.CQ.K .-NMT...GY .YSY..V.HL VLDF.-AEP. HKAQV--L.. TTTPD..EN. E.D------ ----..F.N- R.M.FTEGSE GE------M KSEDVSMRDI  [600]
#AT4G01080_TBL26                 DHKWTV..PK F----.YV.I SG.K-.FLKT TIFHENNVVT G-CH-.CQ.R .-NLTD..GY DYSY..T.NL LRDF.---L. STHKPL.L.. TTTPD..EN. E..T------ ----..Y.N- R.M.F.EG-Q AN------M KTVDDVMRD.  [600]
#AT1G70230_TBL27/AXY4            DERWGNDLER F----.TV.V SV...FL.P AVYYESG.VL G-CH-SC..T S-NCT.V-GF YDV....IR. TL.A.----- AGSGRE.ILT TF.P...E.R P.D------ ----L...N- M.K.YEGKV- ---------L EGLDLDMR.I  [600]
#AT2G40150_TBL28                 E.H-GVNWI. V----.FL.. .SYI-..MNT VS-----IKVL R-GS-FDD.D T-E.D.I-KR PI.YERV.R. L.D..DH.ID .LSTT-..M SM.PL.IKSS D.AN------ -PE-.IR.AL ..T..L.M-S FNVAYGQFSA VGTDYRLFP.  [600]
#AT3G55990_TBL29/XOAT1/ESKIMO1   E.H-GVNW.. V----.FL.. ..YI-..MNT FA----MKVL R-GS-FDK.D T-E.E.I-ER PV.Y.RVMR. ..D...R.ID .LRTT-..A SM.PL.IKSL D.EN------ -PD-.IK.AL ..T..L.M-S MPF------S VGTDYRLFS.  [600]
#AT2G40160_TBL30                 S.H-G.NW.D .-----.YLI. .YI-...RH ST----IKVL KQES-FNK.D SKE.N.I-GI YIVYKQV.S. .TK.L.Q.I. .SQTS-I..S SM.PT.IRSS D.GF------ -NE-.SK.EK .....L.M-S KPI------N VGTNRRLYEI  [600]
#AT1G73140_TBL31                 E.H-.KSWE. V----..L.. ESYV-..M.Q ------PKIN A-T.-----.D TSEVR.Y-N. TT.YKM..E. .AK.FKTKI. SE.QK---.T SM.PT.LWSW E..P------ -GS-D.T.YD .LY..DKR-S .W------G TGSNQEIM.I  [600]
#AT3G11030_TBL32                 N.H-GRHWR. .----.IV.. ..YL-..RTG FK----MKIL .-GS-FDKEK K-RIV.M-ES ED.Y.M..K. MVK..K..MD .L.TR-...A TM.PT.YK.E D.GG------ -EQ-.KN.YN Q.T..QDM-N HW------P SDC..TLM..  [600]
#AT2G40320_TBL33                 N.H-GRHW.. V----.I.I. ..YL-..MTG LK----MNIL Q-GS-FDDKE K-NIV.V-ST ED.Y.MGMKS ML...KN.MD RK.TR-...T SM.PT.AK.I D.GG------ -EP-.QN.YN Q.TL.EDP-S .W------G SDCR.SIM..  [600]
#AT2G38320_TBL34                 E.H-ARHWTN S----.I... .SYL-..RMP H-----IKSL W-GS-FEKLD G-I.K.V-EM VRVYEM..Q. LSQ.L.VH.. .NITK-L..M SM.PT.ERAE E.GG------ -IL-NQN.YG .ASL.DKE-G .T------G RGSDPKMMR.  [600]
#AT5G01620_TBL35                 L.H-ASKWQH .-----.ILI. ..YL-..RQD S-----VKLR W----SS.EK G-SCE.V-KS A.GMEM.MDS ..D..AN..D .N.KR-...V TM.PT.QWSR E..P------ -GS-E.N.YG .KK..EEE-S .W------G SGSDIPTMRM  [600]
#AT3G54260_TBL36                 EDN-ARYWR. V----..L.. DSA.-....S Q-----RWSS W-..-.MD.N K-IFKAM-DP MV.YERG... .AK...I.LD .S.TK-.I.. TV.PRES--- ---------- -----.QM.YN QKH.LPSL-S SS-------. KPHVPQQSR.  [600]
#AT2G34070_TBL37                 EDG-ADAW.N M----.LL.. .SW.-....T GV----Q.Q. W-.F-IRD.. S-LMRDM-DR .D..N.G... ..Q..DQ... VSQTR-...Q .I.PT.YM.R E..E------ -P---RKT.NG QMQ.LTGS-. .P-------G GSLPAA--SI  [600]
#AT1G29050_TBL38                 EGG-ADAW.N M----.L... .SW.-....K G-----Q.Q. W-..-IRD.. S-LVRDM-NR .D..Y.G.S. .A...DQ..D T..TR-...Q .I.PT.YE.R E..E------ -P---RKT.SG QMQ.LGGS-S .P-------S GQ.PSS--G.  [600]
#AT2G42570_TBL39                 KQG--.NMWR. M----.LI. .SW.-....T E-----HIQP W-..-MED.N R-L.KDM-NR .V..Y.GM.. .A...NAY.D .S.TK-...N .V.PT.YE.K D.GE------ -P---MNS.R Q.Q.FYGR-K .P-------G GT.MAW--VI  [600]
#AT2G31110_TBL40                 KQA--.DAWL. M----..LI. .SW.-....T S-----GLQP W-..-MR..N Q-L.KDM-NR .V.YY.G.N. .A..INN.IV .SRTQ-...Q .V.PV.YD.R E..E------ -P--LKS.NG Q.Q.FMGQ-R .P-------G GL.LGW--V.  [600]
#AT3G14850_TBL41                 NDG--.KNWVE M----.TLI. ..W.-..SRR G-----PAQP W-.L-I.I.T -.-.TKDM-DR VA..EI..G. ..K..DTVL. TK.TR-...Q .I.P..YK.V L.GE------ -PA-AKS.VG QK..LLGT-K .P-------G GL.AEV--G.  [600]
#AT1G78710_TBL42                 SRG--.N.WL. S----..AI. ..F-..S.T G-----RA.T W-..-F.T.D K-IVK.M-NR M...KI.... .SK.IDH.ID .S.TR-.YQ .V.PV.LN.. E.GK------ -P--.KT.LG ..V.VQGP-S .P-------G RPNEGE--AI  [600]
#AT2G30900_TBL43                 SSG--SLW.T .----.L.. .SW.-..L.T D-----RKQP W-.A-IMS.N V-TVKDM-DR .V.YE..MM. .AK.IDQ.ID .S.TK-...Q .I.PD.GRAR E.SK------ -QGGK.S.IG ..K..MGS-S .-------A GPHAAE--M.  [600]
#AT5G58600_TBL44/PMR5            SGN-ANAWHD .----.LLI. ....-..S.T G-----SMQ. W-.L-I.S.N S-Y.QDM-DR FV.ME...R. .AY...TH.D RSRTQ-.L.L SI.PT.DNPS D.AASS---- -SSGSKN.YG .....TGT-A .P-------V SSYTDQLRS.  [600]
#AT2G30010_TBL45                 SVDA.NAWRT .----.LL. ....-..S.T G-----SLR. W-EQ-MET.G R-Y.GDM-DR .V.L..G.G. .SS..LRYI. SPLTR-...L SV.PT.YNPN E.T.RSKTST ITQG.KS.YG Q.T.FSGT-. .P-------. SSYVNQ-K..  [600]

#AT5G06700_TBR                  LEKVLR-G-- -------M-- -KTPVTYLNI TRLTDYRKDG HPSVYRKQSL SE---KEKKS PLLYQDCSHW CLPGVPDSWN EILYAELIVK LNQLSQTQRK T*-------- -- [712]
#AT3G12060_TBL1                 ..R...-.-- -------.-- -.........- .........A ...I....K. .A---E.S.. ......... ......... ...F....L.. .D..GGKR.R KALKDHRS*- -- [712]
#AT1G60790_TBL2                 ..YI..DT-- -------.-- -....I.M.. S....F.... ....I..MVYR T.-----.RE AVSH..... ........T.. QL..VS.LKA GLASKW*--- ---------- -- [712]
#AT5G01360_TBL3                 VSS.IK-H-- -------.-- -T.H..VI.. .Q.SE..I.A .T...TETGG KILTAEQRAD .MHHA..I.. ....L..T.. R..L.H.*-- ---------- ---------- -- [712]
#AT5G49340_TBL4                 V.S.IS-E-- -------.-- -....F.M.. .KM.W..T.. ......QPAD PR---GTSPA AGM..... .......... QL...T.L.S HGS.PDKSLG SLL*------ -- [712]
#AT5G20590_TBL5                 VQEAIK-E-- -------.-- -QV..IL..V .K..NF.... ...I.G.TNT DG------.V STRR..... .......V.. HLI..S.LLQ PHS*------ ---------- -- [712]
#AT3G62390_TBL6                 V.D..K-Q-- -------.-- -R....L..V SG.SQ..I.A ...I.GTKPE NR-------R SRAV..... .......T.. HF..LH.LH. R*-------- ---------- -- [712]
#AT1G48880_TBL7                 IKE.VK-N-- -------.-- -TI..SI.DV .SMSAF.S.. .VGLWSDNP. VP-------- -----..... .....I.. ...LFF.FRQ PV.*------ ---------- -- [712]
#AT3G11570_TBL8                 AIQEM.YE-- ---------- -HSK.KF... .Y..EF...A ...R..EPGT P.-------- -DAP..... .......T.. .....Q.LAM NYRTK*---- ---------- -- [712]
#AT5G06230_TBL9                 VIEEM.YR-- ---------- -HSK.KF... .Y..EF.... .I.R..E.GT .V-------- -DVP..... .......T.. .....Q.LSM NYRTK*---- ---------- -- [712]
#AT3G06080_TBL10                .RD..SHNSN RSE------- -TVK.KL... .AMAAQ..... ...L.YLGPH GP-------A ..HR..... .......T.. .LF..LFMKQ EAPS.SKRVE EANSTGNVTM S* [712]
#AT5G19160_TBL11                .QD..S-SSL YYSNISE--- -TVKLKV... .AMAAQ.N.. ...L.YLGLA GP-------A .FHR..... .......... .L...LFLKH EGYS.PRSNN SDTDNFT*-- -- [712]
#AT5G64470_TBL12                IKNE.Q---- ---------- -T.KIKL.DL .H.SEF.A.A .AIWLGKQD AV-------- AIWG...M.. .......T.V D..AELILTN .KTE*----- ---------- -- [712]
#AT2G14530_TBL13                NQHLYN---- -------SLK SRSAFHV.D. ..MSE..A.A ...AAAGGKN- ---------- ---HD..M.. ....LT.T.. DLFV.T.HTI KAL*------ ---------- -- [712]
#AT5G64020_TBL14                V.SAVN---- ---------- -G.RIKI.D. .A.SEL.DEA .I.GSKLKPR KPKKASNVT. TPTIN..L.. ....I..T.. .LFI.QI*-- ---------- ---------- -- [712]
#AT2G37720_TBL15                V.SAVN---- ---------- -G.R.KI.D. .A.SKL.DEA .IAGCKLKPK KA---SNVT. APTFN..L.. ....I..T.. .L.I.Q.*-- ---------- ---------- -- [712]
#AT5G20680_TBL16                AGRAVK---- ---------- -G.G.KL.D. .A.SHI.DE. .I.RFSISAS RG-------- ---V...L.. .....T.. ...F.MI*-- ---------- ---------- -- [712]
#AT5G51640_TBL17/YLS7            QATGFH-RAV ADDKLGNR-- -SKKLKLMD. .EAFG..H.. ..GP..SPDP KKITKRGPDG QPPP...L.. .M..PV.T.. .MVLEIIRRD FEGRQSSPSS *--------- -- [712]
#AT4G25360_TBL18                 QATGYNQA-- VDKVAENL-- -.LKLKLMD. .EAFG..H.. ..GPF.SPDP NKITKRGPDG RPPP...L.. .M..PV.T.. .MVLEIIRRD RKSS.T*--- ---------- -- [712]
#AT5G15900_TBL19                 QLEEFW-RAE -EEA-KKK-- -GKRLRL.DT .QAMWL.P.. ...R.GHIPE AN-------- VT..N..V.. ....PI.NL. DF.L.M.RRE EDKGFLA.VR KMLS*----- ---------- -- [712]
#AT3G02440_TBL20                 --------- ---------- --------- --------- --------- --------- ---------- ---------- ---------- ---------- -- [712]
#AT5G15890_TBL21                 QREEF.-AAE -EDGMKKS-- -GLRLKLMDT .QAMLL.P.. ...GR.GHLQN PN-------- VT.RN..I.. ....PI.TL. D..LQMMKTD N*-------- ---------- -- [712]
#AT3G28150_TBL22/AXY4-L          QIEQ.E-.IT -KR--GNK-- -AKKFAV.DV ..VMQM.P.. ...NG.WGNKW MK-G------ ----N..V.. ....PI.A.. DF.M.IIRQL R*-------- ---------- -- [712]
#AT4G11090_TBL23                 EINQFE-R-- ---VVTE.GQ ESENLKL.DF AGMLLT.P.. ...GP..EFRP FD----KD.N ATVQN..L.. ....PI.HL. DVILEIIVNG RTGK*----- ---------- -- [712]
#AT4G23790_TBL24                 EIDQFK-RAV REKTNQD--- -GGNLKL.DF ..MLLT.P.. ...GE..QFRP FD----KD.N AKVQN..L.. ....PF.YL. DVILETIVNG *--------- ---------- -- [712]
#AT1G01430_TBL25                 ELEEFY-KTT -TTQQEGS-- -NSNIVL.DT .SMSLL.P.. ...GP..YPNP FA-GL.N.EL NQVQN..L.. ....PI.... DLMVEVMLNR ER.RRE*--- ---------- -- [712]
#AT4G01080_TBL26                 ELE.FQ-K-- -FGKGFGL-- -GSNIRL.DT .GMSLL.P.. ...GP..HPNP FA---GV.NK SNVQN..L.. ....PI.... DVMVETTLNR ERE.YDLTG* ---------- -- [712]
#AT1G70230_TBL27/AXY4            EIEEYT---- -------AAA AEVRLEV.DV .AMSVL.P.. ...GP.MYADP FK-----NGVP ERIPN..L.. ....PV.T.. .MIEM.RRW KV*------- ---------- -- [712]
#AT2G40150_TBL28                 A.N.TQ-S-- -------L-- -.V.IHF... .A.SE....A .T...TIKQG KLLTEQQND .ANFA..I.. ....L..T.. .F..THI.SR R*-------- ---------- -- [712]
#AT3G55990_TBL29/XOAT1/ESKIMO1   A.N.TH-S-- -------L-- -NV..YF... .K.SE....A .T..HTIRQG KMLTPEQQAD .NT.A..I.. ....L..T.. .F..TRI.SR S*-------- ---------- -- [712]
#AT2G40160_TBL30                 ALNATK-S-- -------T-- -.V.IHF.... .TMSE..... .T.F.GSING KLMTPEQ.LD .RTFA..Y.. ....L..... .L.SLF.LHR ER.KVTGRDE RATAK*----- -- [712]
#AT1G73140_TBL31                 VGD..S-R-- -------V-- -GEN..F... .Q.SE..... .TT..GERRG KLLTKEQRAD .KN.G..I.. .......T.. .....Y.LRS HRNFF*---- ---------- -- [712]
#AT3G11030_TBL32                 IGEE.DQR-- -------A-- -EF...V... .Q.SG....A .T.I.K..WS PL-TKEQLAN .AS.S..I.. ....LQ.T.. .LFF.K.FYP *--------- ---------- -- [712]
#AT2G40320_TBL33                 IGE.FG-R-- -------S-- -...I.L... .QMSN....A .T.I.K..WS PL-TAEQLEN .TS.A..V.. ....LQ.T.. .L.F.K.FYT *--------- ---------- -- [712]
#AT2G38320_TBL34                 ..N..D-.-- -------LKN RGLNMQMI.. .Q.SE...E. ...I....WG TV-KEN.ISN .SSNA..I.. .......V.. .L...YILDH HSS*------ ---------- -- [712]
#AT5G01620_TBL35                 VKR..E-R-- -------L-- -GPK.SVI.. .Q.SE..... .......FWE PL-NEDRL.N .AS.S..T.. .V.....V.. QL.FHF.*-- ---------- ---------- -- [712]
#AT3G54260_TBL36                 .N....-T-- --------.-- -.YR.YLYD. .TMSA..R.. ....FKRAMH E.--E.HHRI AGPSS..... .......I.. .M.SSILTN AV*------- ---------- -- [712]
#AT2G34070_TBL37                 VSR..S-T-- --------.-- -R...YL.D. .T.SQL...A ...T.GGDGG ---------- ----T..... ....L..T.. QL...A.SM* ---------- ---------- -- [712]
#AT1G29050_TBL38                 VS..S-S-- --------.-- -.K...L.D. .T.SQL...A ...S.GGDGG ---------- ----T..... ....L..T.. QL...A.TM* ---------- ---------- -- [712]
#AT2G42570_TBL39                 .N..M.-R-- -------L-- -.K..HW.D. .G.SQL...A ...AFSGNHP GN-------- -----..... ....L..T.. LLF.ST.FSS *--------- ---------- -- [712]
#AT2G31110_TBL40                 VN...S-R-- -------I-- -RK..HL.DL .T.SE....A ...L.NGI.K DL-------- -----..... ....L..T.. LL..SS.TS* ---------- ---------- -- [712]
#AT3G14850_TBL41                 .KRA.G-K-- -------I-- -SK...L.D. .M.SLL...A .....GL--G GR-------- -NSSG..... ..S....T.. ....NYMVE* ---------- ---------- -- [712]
#AT1G78710_TBL42                 VKS.IG-R-- --------.-- -AK..EL.DV .AM.EM.... ...I.AGGGD RL-------- ----N..... .......A.. QL..TA.LSH *--------- ---------- -- [712]
#AT2G30900_TBL43                 VA..IK-T-- --------.-- -.NQARLMDV .LMSQL.... .....GFGGH RM-------- ----A..... ..S..... QL..S..FHS *--------- ---------- -- [712]
#AT5G58600_TBL44/PMR5            IVE..H-.-- --------.-- -HN.AFL.D. .L.SSL.... .....SGLIS GS----QRSR .DQSA..... ....L..T.. QL..TL..Y* ---------- ---------- -- [712]
#AT2G30010_TBL45                 IDD.VK-E-- -------.-- -.SH.SLMD. .M.SAL.V.. ...I.SGDLN PS----L.RN .DRSS..... ....L..T.. QLF..A.LY* ---------- ---------- -- [712]
```

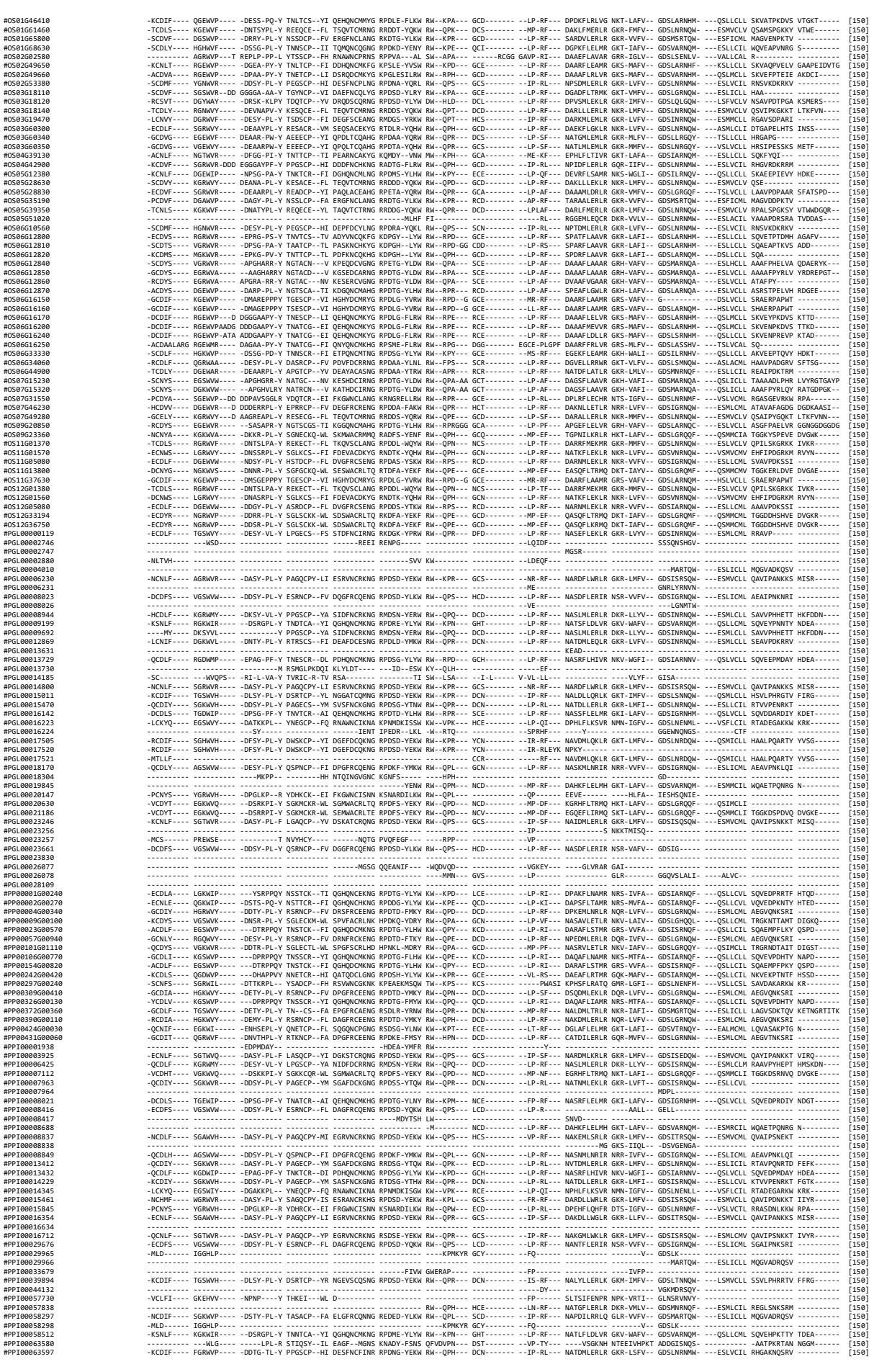
Figure S4: Alignment of ESK1 clade from 11 plant species.
This alignment was used to build the Maximum likelihood tree shown in this annex (Figure S5). The variable N-terminal region was manually removed from this alignment.

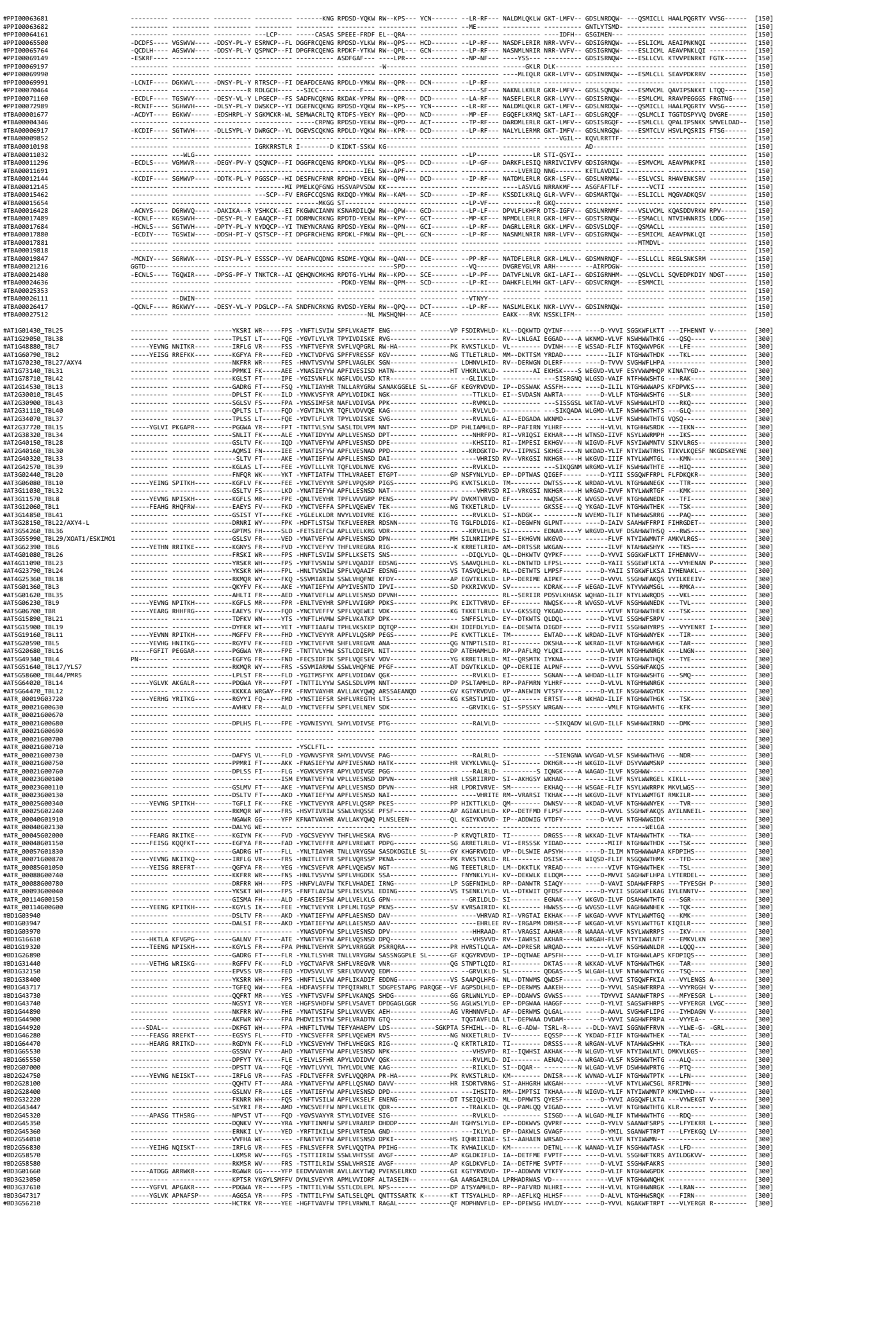
Key:

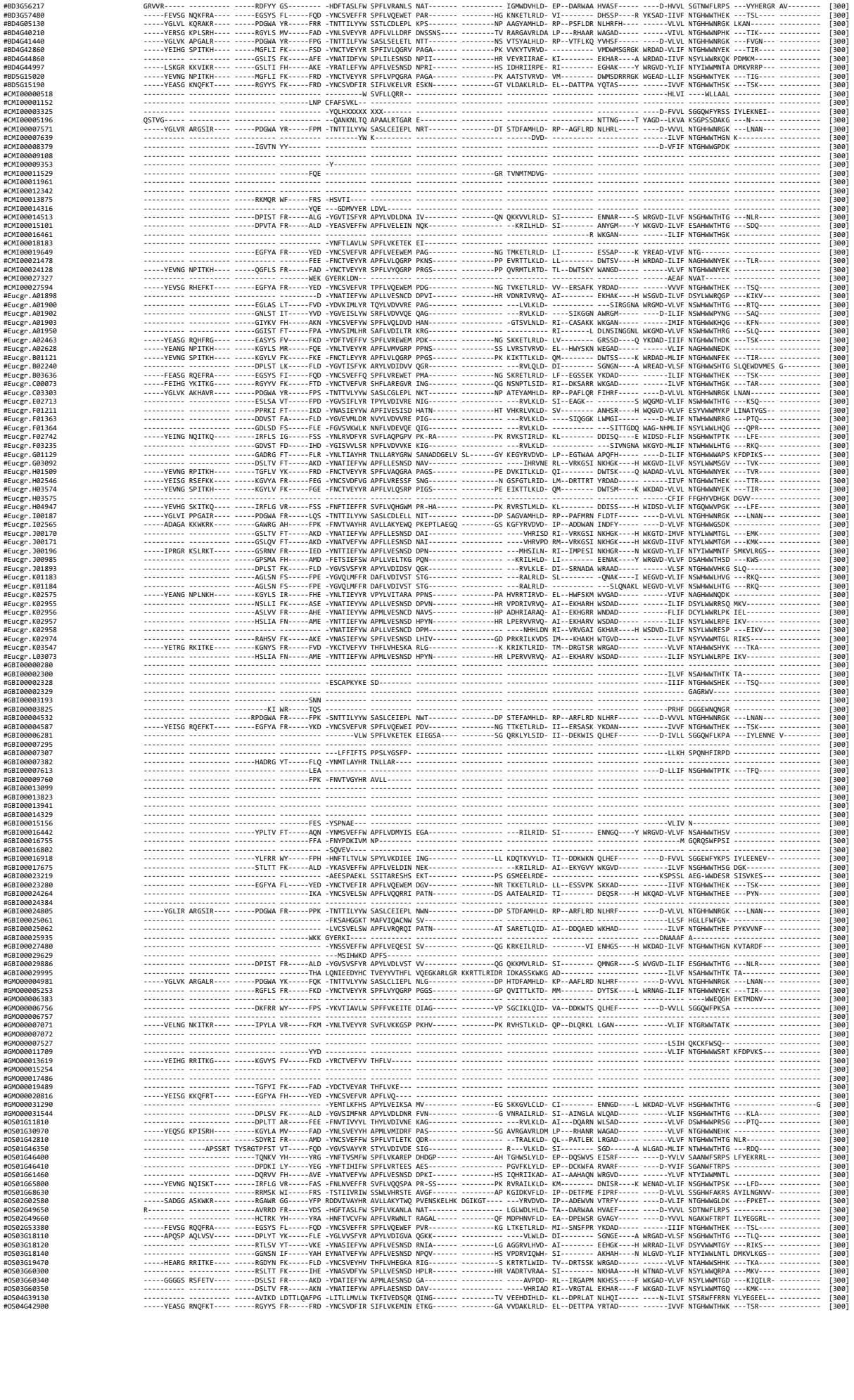
AT *Arabidopsis thaliana*
ATR *Amborella trichopoda*
BD *Brachypodium distachyon*
CMI *Cycas micholitzii*
Eucgr. *Eucalyptus grandis*
BGI *Ginkgo biloba*
GMO *Gnetum montanum*
OS *Oryza sativa*
PGL *Picea glauca*
PP *Physcomitrella patens*
TBA *Taxus baccata*

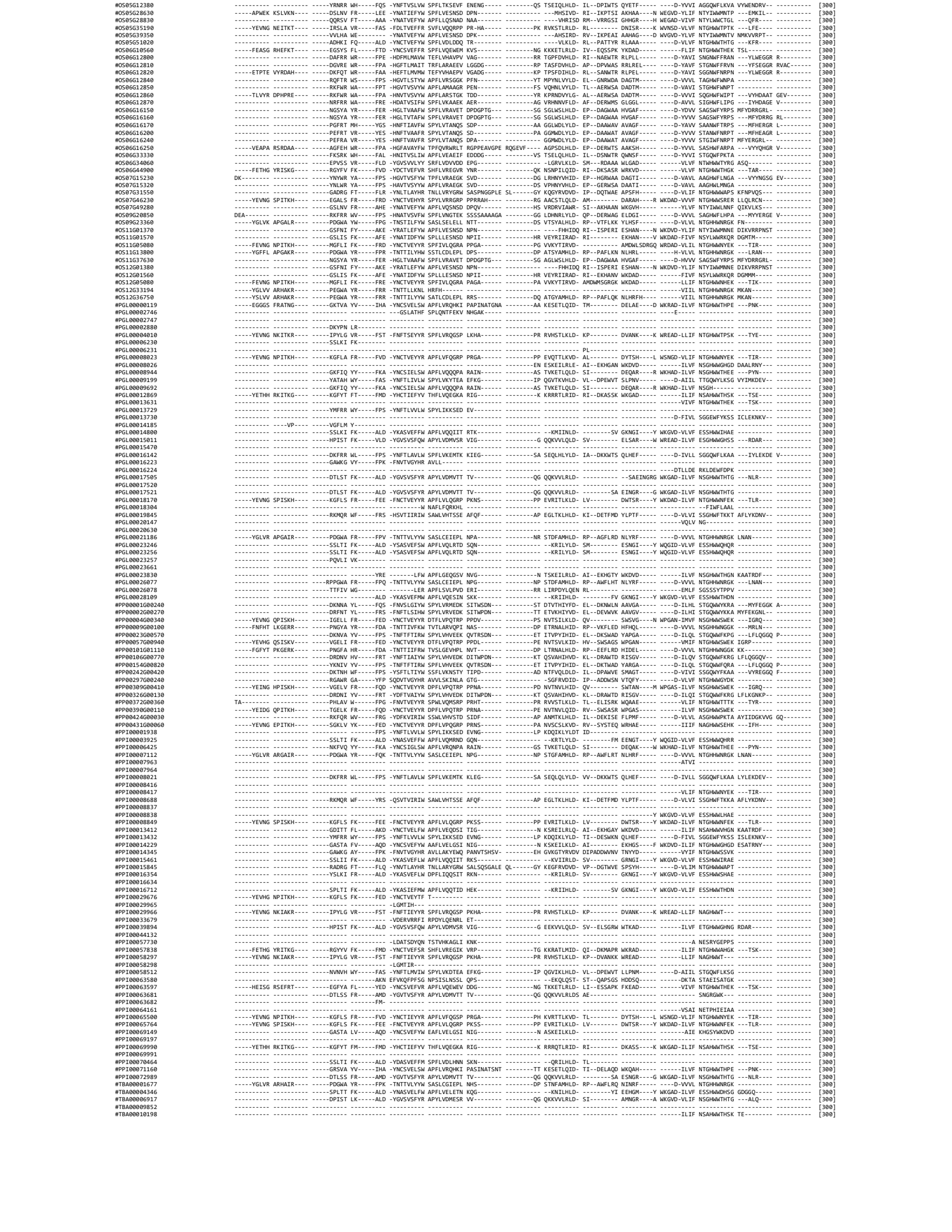


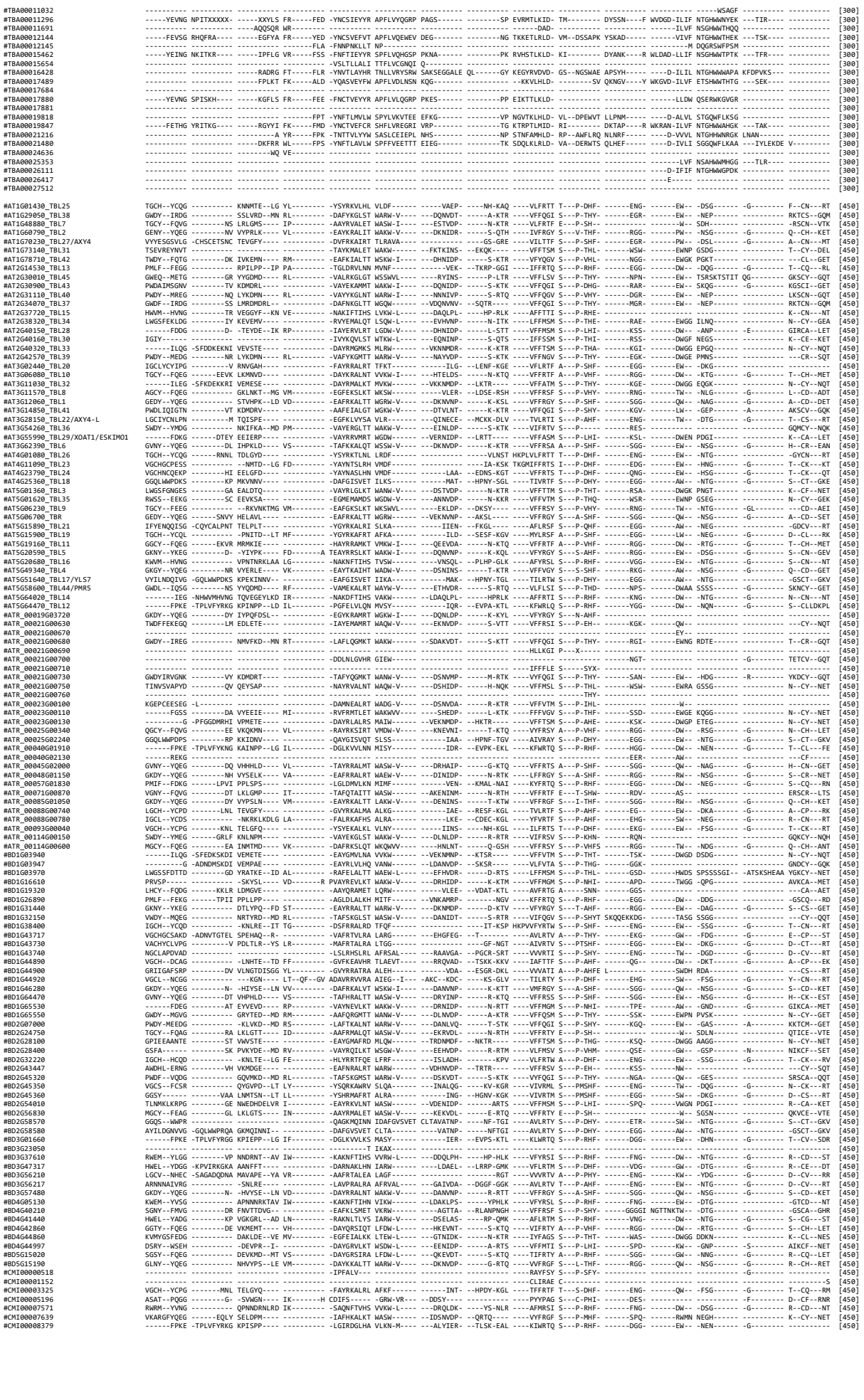


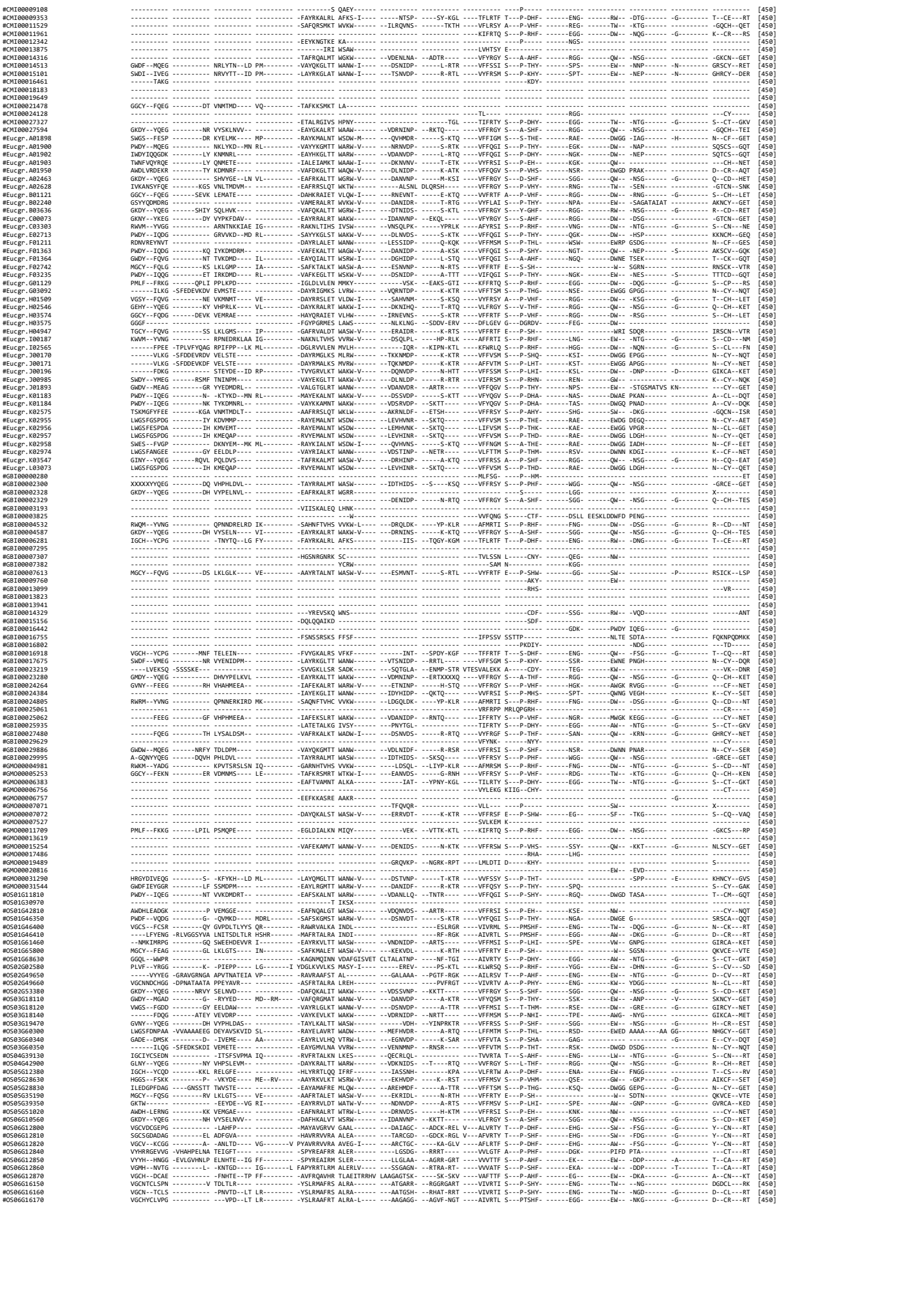


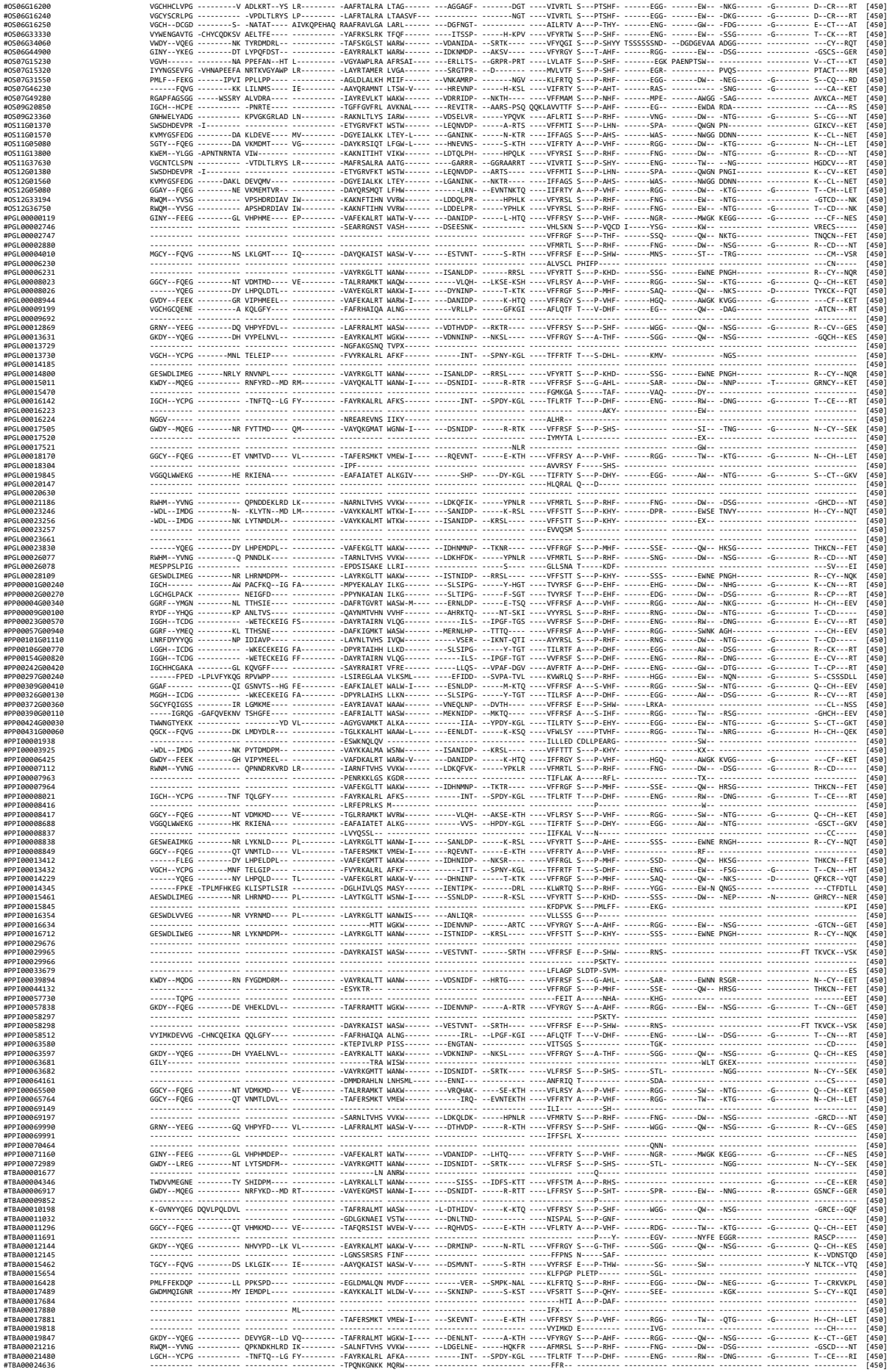


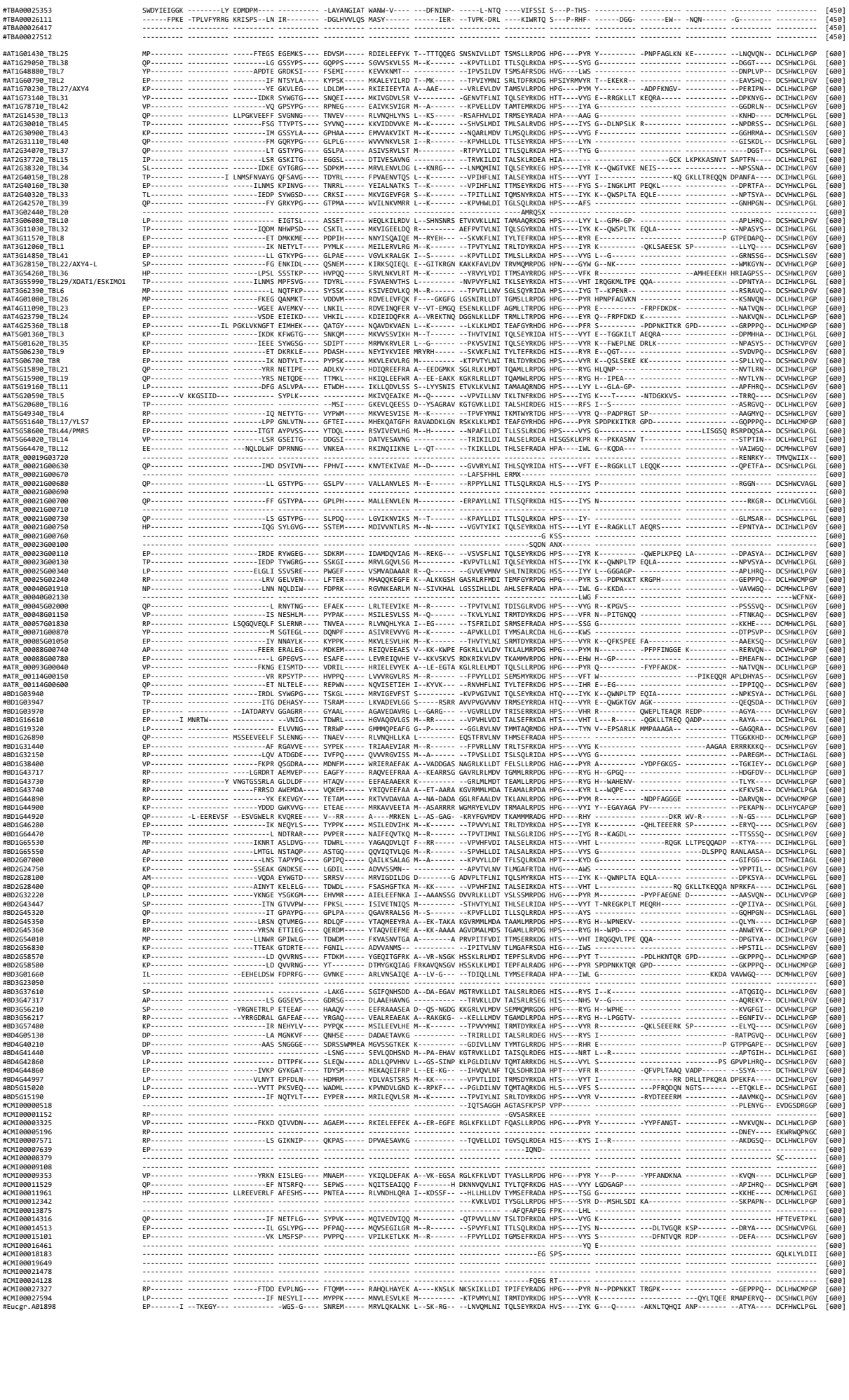


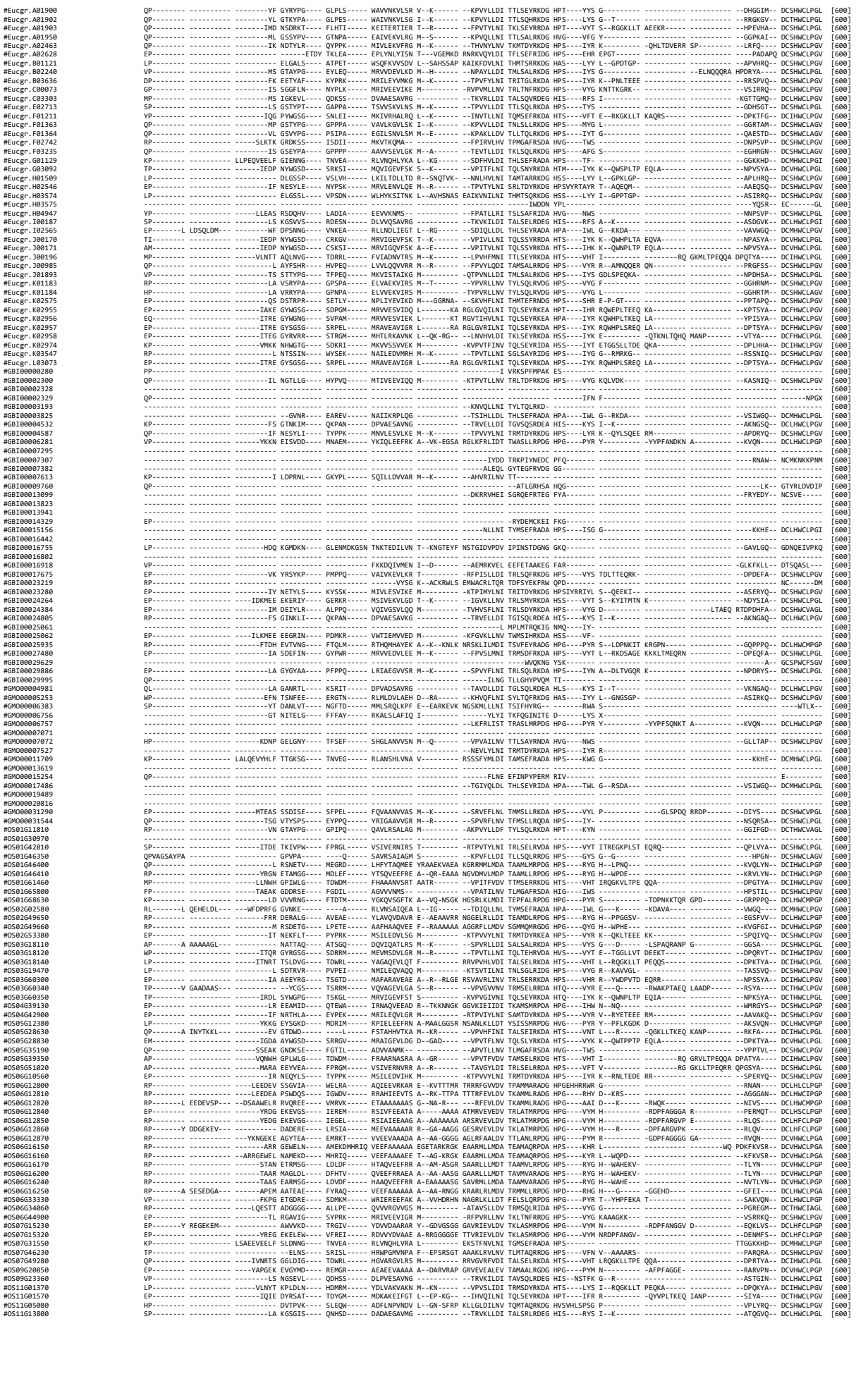


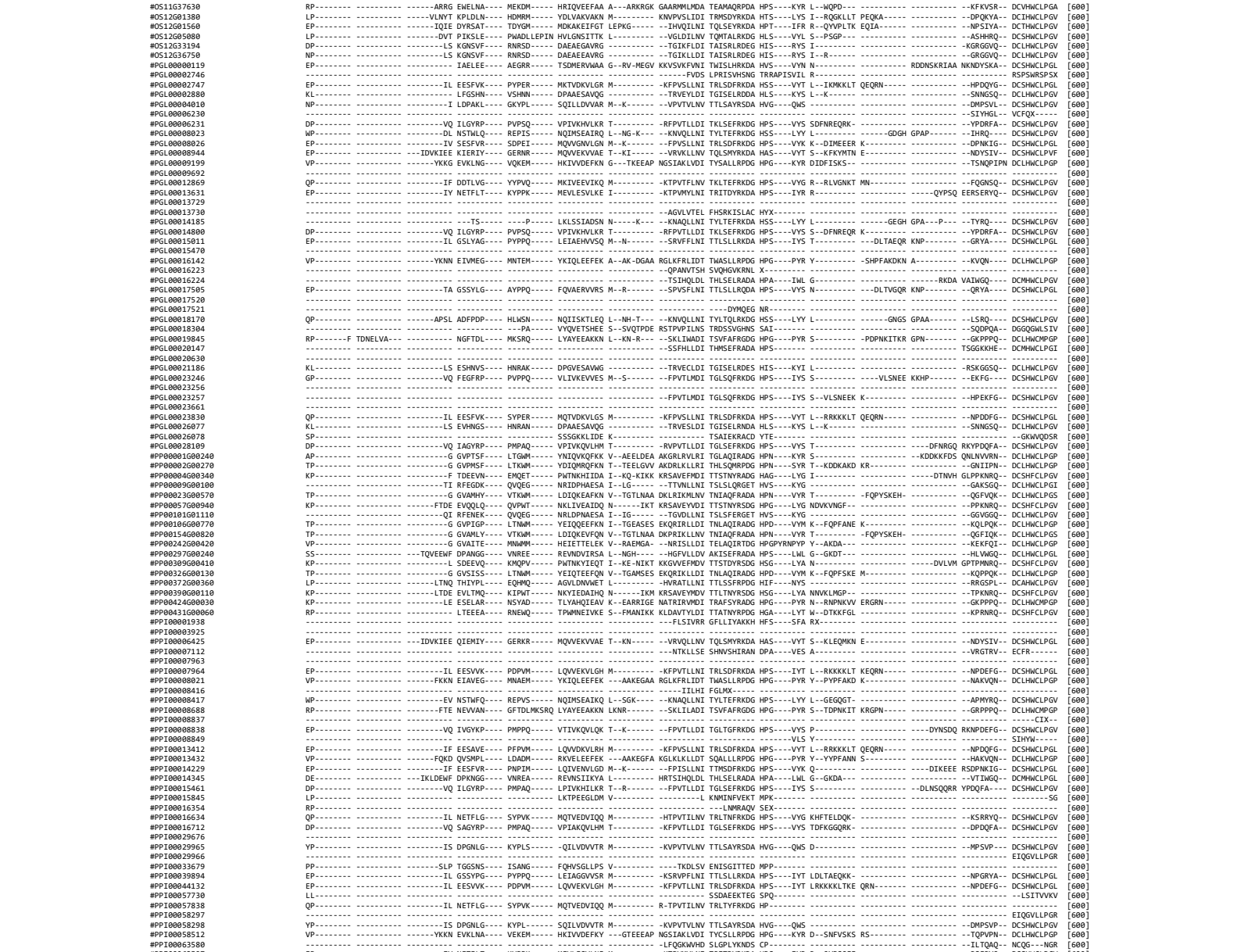








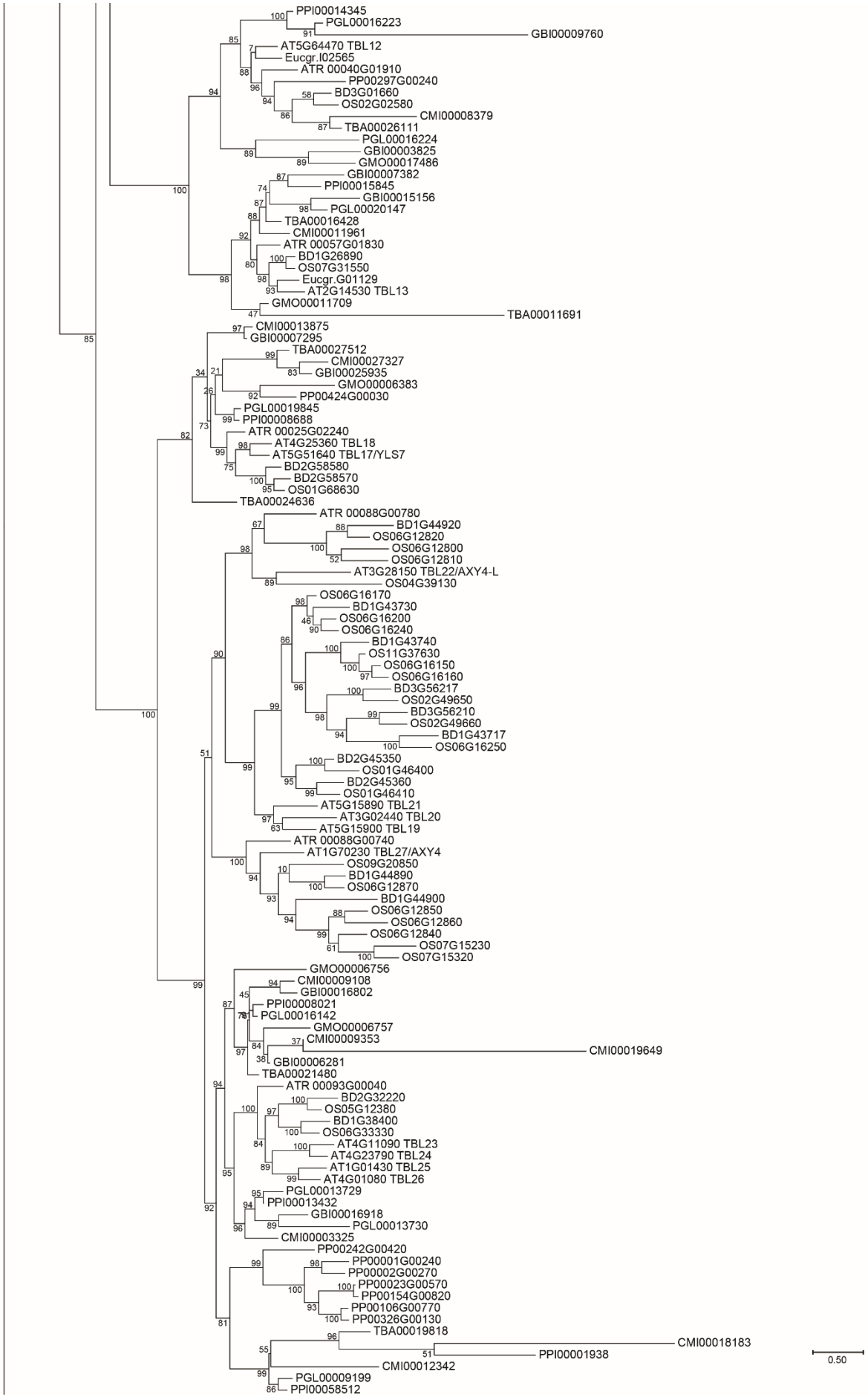


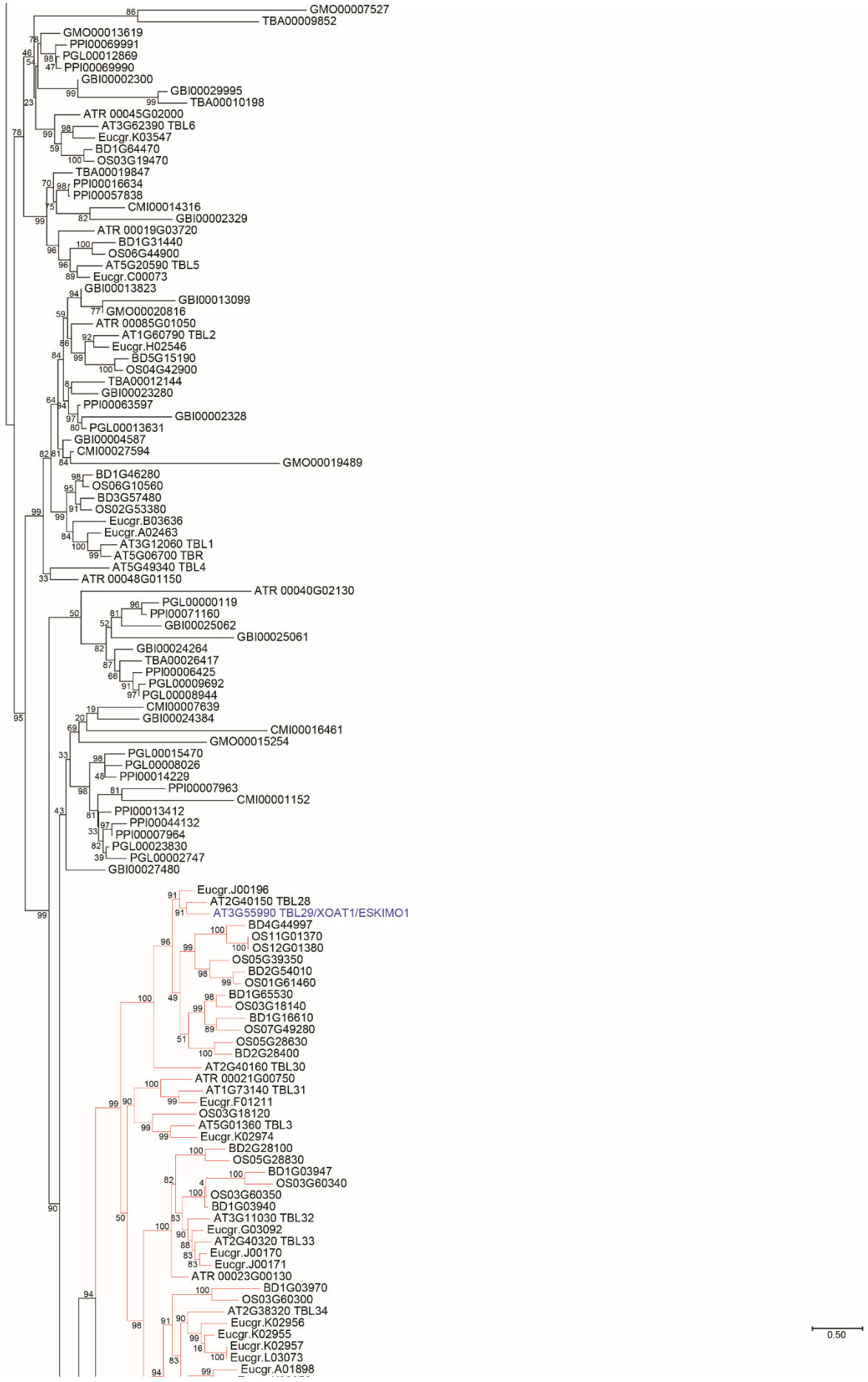


```
#AT3G28150_TBL22/AXY4-L          IDAWNDFLMA IIRQLRX--- ---------- ---------- --------- [649]
#AT3G54260_TBL36                 PDIWNEMLSS IILTNAVX-- ---------- ---------- --------- [649]
#AT3G55990_TBL29/XOAT1/ESKIMO1   PDTWNEFLYT RIISRSX--- ---------- ---------- --------- [649]
#AT3G62390_TBL6                  PDTWNHFLYL HLLHKRX--- ---------- ---------- --------- [649]
#AT4G01080_TBL26                 IDSWNDVMVE TTLNRERELY DLTGX----- ---------- --------- [649]
#AT4G11090_TBL23                 IDHLNDVLIE IIVNGRTGKX ---------- ---------- --------- [649]
#AT4G23790_TBL24                 FDYLNDVLIE TIVNGX---- ---------- ---------- --------- [649]
#AT4G25360_TBL18                 VDTWNEMVLE LIRRDRKSSS TX-------- ---------- --------- [649]
#AT5G01360_TBL3                  PDTWNRILLA HLX------- ---------- ---------- --------- [649]
#AT5G01620_TBL35                 PDWWNQLLFH FLX------- ---------- ---------- --------- [649]
#AT5G06230_TBL9                  PDTWNEILYA QLLSMNYRTK X--------- ---------- --------- [649]
#AT5G06700_TBR                   PDSWNEILYA ELIVKLNQLS QTQRKTX--- ---------- --------- [649]
#AT5G15890_TBL21                 IDTLNDILLQ MMKTDNX--- ---------- ---------- --------- [649]
#AT5G15900_TBL19                 IDNLNDFLLA MLKREEDKGF LAQVRKMLSX ---------- --------- [649]
#AT5G19160_TBL11                 PDSWNELLYA LFLKHEGYSS PRSNNSDTDN FTX------- ---------- [649]
#AT5G20590_TBL5                  PDWNNHLIYA SLLLQ----- PHSX------ ---------- --------- [649]
#AT5G20680_TBL16                 PDTWNEILFA MIX------- ---------- ---------- --------- [649]
#AT5G49340_TBL4                  PDSWNQLLYA TLLVSHGSLP DKSLGSLLX- ---------- --------- [649]
#AT5G51640_TBL17/YLS7            VDTWNEMVLE IIRRDFEGRQ SSPSSX--- ---------- --------- [649]
#AT5G58600_TBL44/PMR5            PDTWNQLLYT LLIYX----- ---------- ---------- --------- [649]
#AT5G64020_TBL14                 PDTWNELFIA QIX------- ---------- ---------- --------- [649]
#AT5G64470_TBL12                 PDTWVDILAE LILTNLKTEX ---------- ---------- --------- [649]
#ATR_00019G03720                 ---------- ---------- ---------- ---------- --------- [649]
#ATR_00021G00630                 PDTWNELLYA SLLLGTPRQV LX-------- ---------- --------- [649]
#ATR_00021G00670                 ---------- ---------- ---------- ---------- --------- [649]
#ATR_00021G00680                 PDTWNLLLYA ALFFX----- ---------- ---------- --------- [649]
#ATR_00021G00690                 ---------- ---------- ---------- ---------- --------- [649]
#ATR_00021G00700                 PDTWNLLLYT ALFFX----- ---------- ---------- --------- [649]
#ATR_00021G00710                 ---------- ---------- ---------- ---------- --------- [649]
#ATR_00021G00730                 PDTWNQLLYT ALYFX----- ---------- ---------- --------- [649]
#ATR_00021G00750                 PDTWNEILYT YLVVHKX--- ---------- ---------- --------- [649]
#ATR_00021G00760                 ---------- ---------- ---------- ---------- --------- [649]
#ATR_00023G00100                 ---------- ---------- ---------- ---------- --------- [649]
#ATR_00023G00110                 PDTWNELLYA SLFSHQX--- ---------- ---------- --------- [649]
#ATR_00023G00130                 QDTWNELLYA KLFFPX---- ---------- ---------- --------- [649]
#ATR_00025G00340                 PDSWNHLLYA FILHHDMHHR TSSX------ ---------- --------- [649]
#ATR_00025G02240                 IDTWNELVFE IIRREFEGIR X--------- ---------- --------- [649]
#ATR_00040G01910                 PDTWVDILST LIHERLVKKT GPAIX----- ---------- --------- [649]
#ATR_00040G02130                 ---------- ---------- ---------- ---------- --------- [649]
#ATR_00045G02000                 PDTWNELLYF LLVSKRREVS TKX------- ---------- --------- [649]
#ATR_00048G01150                 PDAWNELLYA TLLISGRGTW GKKKNQHX-- ---------- --------- [649]
#ATR_00057G01830                 TDFWNDLLVV NLITNX---- ---------- ---------- --------- [649]
#ATR_00071G00870                 PDAWNELLLT LTAPKNGTQP AX-------- ---------- --------- [649]
#ATR_00085G01050                 PDTWNELLYA SLLIKGFX-- ---------- ---------- --------- [649]
#ATR_00088G00740                 VDTWNRLLLE MVTRYQKSRK PHYPANVPX- ---------- --------- [649]
#ATR_00088G00780                 VDAWNELLME MVARERSKEX ---------- ---------- --------- [649]
#ATR_00093G00040                 IDSWNDIVMK IVLDERRSRX ---------- ---------- --------- [649]
#ATR_00114G00150                 PDAWNEILYA SIMX------ ---------- ---------- --------- [649]
#ATR_00114G00600                 PDTWNELLYA NLLSNGYRTD TNNTRKSX-- ---------- --------- [649]
#BD1G03940                       QDTWNELLYS KLFFPX---- ---------- ---------- --------- [649]
#BD1G03947                       PDAWNELLYW KLFFPAADDQ ALX------- ---------- --------- [649]
#BD1G03970                       PDWNHLLYA HIVAX----- ---------- ---------- --------- [649]
#BD1G16610                       PDTWNHFLYA RIIADSRSSS STTPASSLDI SATSSX---- ---------- [649]
#BD1G19320                       PDAWNELLYA MIINKFSX-- ---------- ---------- --------- [649]
#BD1G26890                       TDTWNDLLAT NLAAIESX-- ---------- ---------- --------- [649]
#BD1G31440                       PDWWNELIYA SFVMEPSPSS WNRRX----- ---------- --------- [649]
#BD1G32150                       PDAWNHILYA MLLX------ ---------- ---------- --------- [649]
#BD1G38400                       IDTWNDIIMQ MLGNDLKPPP HPRAX----- ---------- --------- [649]
#BD1G43717                       IDWNNELLQ ILDGX------ ---------- ---------- --------- [649]
#BD1G43730                       IDWWNEMLLA MILRSSVVRX ---------- ---------- --------- [649]
#BD1G43740                       MDACNDMLLH MLIGX----- ---------- ---------- --------- [649]
#BD1G44890                       IDTFNEILLQ TVVGX----- ---------- ---------- --------- [649]
#BD1G44900                       VDTFNDVLAR MIAAARGX-- ---------- ---------- --------- [649]
#BD1G44920                       VDWWNGVLLQ RLAEIDPAPL PIDRGEVVNX ---------- --------- [649]
#BD1G46280                       PDSWNELVYA QLLIRQHQML QQX------- ---------- --------- [649]
#BD1G64470                       PDTWNELLFY HLVSSQEIGV RSX------- ---------- --------- [649]
#BD1G65530                       PDTWNEFLYA RIVSSPSTTT EQX------- ---------- --------- [649]
#BD1G65550                       PDTWNQLFYT LLFYQX---- ---------- ---------- --------- [649]
#BD2G07000                       PDTWNVLFYA ALLGQSX--- ---------- ---------- --------- [649]
#BD2G24750                       PDAWNELVFS YLSTNGWRKT AGX------- ---------- --------- [649]
#BD2G28100                       QDTWNELLYT KLFYPX---- ---------- ---------- --------- [649]
#BD2G28400                       PDTWNEFIYG HIVSSPLQQQ IEDQSQRX-- ---------- --------- [649]
#BD2G32220                       IDAWNDLIMK MVLNX----- ---------- ---------- --------- [649]
#BD2G43447                       PDTWNVLLLA SLMRSPADVH LVRKDKX--- ---------- --------- [649]
#BD2G45320                       PDTWNQILYA SLLAX----- ---------- ---------- --------- [649]
#BD2G45350                       IDIWNDLLFQ MILVX----- ---------- ---------- --------- [649]
#BD2G45360                       VDAWSDMLFH MMSDX----- ---------- ---------- --------- [649]
#BD2G54010                       PDSWNSLLYT RIMSRPETGI ATTRX----- ---------- --------- [649]
#BD2G56830                       PDTWNELVFS HLLTNGWRKM AGX------- ---------- --------- [649]
#BD2G58570                       VDTWNEMLFE TIQSDLKRDR SX-------- ---------- --------- [649]
#BD2G58580                       VDTWNEMLFE TIRREFGGDG SX-------- ---------- --------- [649]
#BD3G01660                       PDTWVDILAA RISHYFKQGK GX-------- ---------- --------- [649]
#BD3G23050                       ---------- ---------- ---------- ---------- --------- [649]
#BD3G37610                       PDTWNEILAA QLX------- ---------- ---------- --------- [649]
#BD3G47317                       PDLWNELLFA QIX------- ---------- ---------- --------- [649]
#BD3G56210                       VDAWNELLLH LLSGX----- ---------- ---------- --------- [649]
#BD3G56217                       IDWSSELFLQ MLVHQX---- ---------- ---------- --------- [649]
#BD3G57480                       PDSWNELLYA QILVKQHQMI HQX------- ---------- --------- [649]
#BD4G05130                       PDTWNEILAT QLX------- ---------- ---------- --------- [649]
#BD4G40210                       PDAWNQLLYA QLLSMGYDAR KKRSNKSTX- ---------- --------- [649]
#BD4G41440                       PDMWNELLFA QLX------- ---------- ---------- --------- [649]
#BD4G42860                       PDTWNELLYA VFMKRQAMMD QNVSLAGPRT LSTGX---- ---------- [649]
#BD4G44860                       PDVWNSFLYS YLVHKX---- ---------- ---------- --------- [649]
#BD4G44997                       PDWWNTVLFT RIFSKSSPPP LPSLPPQX-- ---------- --------- [649]
#BD5G15020                       PDTWNELLYA VFMRRQMISD QNVPLSAX-- ---------- --------- [649]
#BD5G15190                       PDTWNELLYA SLLQAGKGSW RLX------- ---------- --------- [649]
#CMI00000518                     EDGTVAVVRS G--------- ---------- ---------- --------- [649]
#CMI00001152                     ---------- ---------- ---------- ---------- --------- [649]
#CMI00003325                     IDTWSELMLE VLKHVX---- ---------- ---------- --------- [649]
#CMI00005196                     NIPR------ ---------- ---------- ---------- --------- [649]
#CMI00007571                     PDTWNELLFA QLLFPQRKKQ PAHFX----- ---------- --------- [649]
#CMI00007639                     ---------- ---------- ---------- ---------- --------- [649]
#CMI00008379                     ---------- ---------- ---------- ---------- --------- [649]
#CMI00009108                     RDIYND---- ---------- ---------- ---------- --------- [649]
#CMI00009353                     IDTWSEFLLE LLKHEX---- ---------- ---------- --------- [649]
#CMI00011529                     PDTWNELLYS YLLARDNEYS STGVIRRX-- ---------- --------- [649]
#CMI00011961                     TDYWNDLLMV TLFNKKMKAF X--------- ---------- --------- [649]
#CMI00012342                     IDTWNEILLQ MIREAX---- ---------- ---------- --------- [649]
#CMI00013875                     -DTVDE---- ---------- ---------- ---------- --------- [649]
#CMI00014316                     LKYQDCXXXQ SLVSSRCSRC MEX------- ---------- --------- [649]
#CMI00014513                     PDTWNELLYA ALLFX----- ---------- ---------- --------- [649]
#CMI00015101                     PDTWNQLLYA SLITRGLGKG X--------- ---------- --------- [649]
#CMI00016461                     ---------- ---------- ---------- ---------- --------- [649]
#CMI00018183                     DEKW------ ---------- ---------- ---------- --------- [649]
#CMI00019649                     ---------- ---------- ---------- ---------- --------- [649]
#CMI00021478                     ---------- ---------- ---------- ---------- --------- [649]
#CMI00024128                     ---------- ---------- ---------- ---------- --------- [649]
#CMI00027327                     IDTWNEFVFE IVRRHLSGID DMX------- ---------- --------- [649]
#CMI00027594                     PDAWNELLYA SLLIRGQGVW ATX------- ---------- --------- [649]
#Eucgr.A01898                    PDVWNELLYA YILQNX---- ---------- ---------- --------- [649]
#Eucgr.A01900                    PDTWNELLYA ALFSX----- ---------- ---------- --------- [649]
#Eucgr.A01902                    PDTWNELFYA VLFSX----- ---------- ---------- --------- [649]
#Eucgr.A01903                    PDTWNRLLYA SLVLDGTRGN SSSSHLAVX- ---------- --------- [649]
#Eucgr.A01950                    PDTWNLLLYA SLIQNX---- ---------- ---------- --------- [649]
#Eucgr.A02463                    PDAWNELLFA ELVKLNQKX ---------- ---------- --------- [649]
#Eucgr.A02628                    PDTWNELLYA HLLSVRFHEM X--------- ---------- --------- [649]
#Eucgr.B01121                    PDSWNVLLYA LFLKHDSNRD QX-------- ---------- --------- [649]
#Eucgr.B02240                    PDTWNQLFYT ALFYX----- ---------- ---------- --------- [649]
#Eucgr.B03636                    PDTWNELLYS QILIKEIEKG WQENSX---- ---------- --------- [649]
#Eucgr.C00073                    PDSWNELIYA SLVFRQTIST SYTX------ ---------- --------- [649]
#Eucgr.C03303                    PDTWNELLFA QIX------- ---------- ---------- --------- [649]
#Eucgr.E02713                    PDTWNQLLYA AIVMX----- ---------- ---------- --------- [649]
#Eucgr.F01211                    PDAWNEILYA YLKKDHRTIL X--------- ---------- --------- [649]
#Eucgr.F01363                    PDAWNELLYA LILX------ ---------- ---------- --------- [649]
#Eucgr.F01364                    PDTCNELLYT TLLKINPNRT SRDAPAPSPQ PAAPPAPSPP PAAPPAPSP [649]
#Eucgr.F02742                    PDMWNEILFS YMLSKNGELF QX-------- ---------- --------- [649]
#Eucgr.F03235                    PDTWNQLLYT ILINRGRRX- ---------- ---------- --------- [649]
#Eucgr.G01129                    TDTWNDIFVM LLTQIRRSRX ---------- ---------- --------- [649]
#Eucgr.G03092                    QDTWNELLFA KLFYPX---- ---------- ---------- --------- [649]
#Eucgr.H01509                    PDWWNELLYA FFLKGDSVRS RNLTNASX-- ---------- --------- [649]
#Eucgr.H02546                    PDTWNELLYA SLLKMGWGSP TX-------- ---------- --------- [649]
#Eucgr.H03574                    PDSWNELLFA LYLKHESSQQ AHGAQKX--- ---------- --------- [649]
#Eucgr.H03575                    MEEEGRIFVV IFDRNLLEY AX-------- ---------- --------- [649]
#Eucgr.H04947                    PDMWNEILLY YLLTNHEX-- ---------- ---------- --------- [649]
#Eucgr.I00187                    PDTWNEILCA QVX------- ---------- ---------- --------- [649]
#Eucgr.I02565                    PDTWADILSQ LIRHGLERX- ---------- ---------- --------- [649]
#Eucgr.J00170                    QDTWNELLFT KLFYTX---- ---------- ---------- --------- [649]
#Eucgr.J00171                    HDTWNELLFT RLFYPX---- ---------- ---------- --------- [649]
#Eucgr.J00196                    PDTWNEFIYT RILQHSX--- ---------- ---------- --------- [649]
#Eucgr.J00985                    PDTWNEMLTA LLX------ ---------- ---------- --------- [649]
#Eucgr.J01893                    PDTWNQLFYT SLLFX----- ---------- ---------- --------- [649]
#Eucgr.K01183                    PDTWNLILYR SLVANNFX-- ---------- ---------- --------- [649]
#Eucgr.K01184                    PDTWNLLLYQ SLVAHNFX-- ---------- ---------- --------- [649]
#Eucgr.K02575                    PDTWNELLYA QLVSRRYVCE X--------- ---------- --------- [649]
#Eucgr.K02955                    PDVWNELLYA HIFQNRTQRL ENRVEGLX-- ---------- --------- [649]
#Eucgr.K02956                    PDVWNELLYT YIFRYIHTX- ---------- ---------- --------- [649]
#Eucgr.K02957                    SDVWNELLYA YILRNRSPCX ---------- ---------- --------- [649]
#Eucgr.K02958                    PDVWNELLYA YILHNX---- ---------- ---------- --------- [649]
#Eucgr.K02974                    PDTWNQILYA HLX------- ---------- ---------- --------- [649]
#Eucgr.K03547                    PDTWNELLYM HLQTKLRHSV QX-------- ---------- --------- [649]
#Eucgr.L03073                    SDVWNELLYA YILRNRSPCX ---------- ---------- --------- [649]
#GBI00000280                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00002300                     PD--------- ---------- ---------- ---------- --------- [649]
#GBI00002328                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00002329                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00003193                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00003825                     PDTWVDVLSA VIMGHLNAKA RCLSSADKX- ---------- --------- [649]
#GBI00004532                     PDTWNELLFA QLLSPRQTNA LSQIX----- ---------- --------- [649]
#GBI00004587                     PDTWNELLYA SLLSRGIGVW RTX------- ---------- --------- [649]
#GBI00006281                     IDTWNEFLLE MLKHEX---- ---------- ---------- --------- [649]
#GBI00007295                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00007307                     EKIF------ ---------- ---------- ---------- --------- [649]
#GBI00007382                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00007613                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00009760                     ADDWIN---- ---------- ---------- ---------- --------- [649]
#GBI00013099                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00013823                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00013941                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00014329                     ---WNC---- ---------- ---------- ---------- --------- [649]
#GBI00015156                     TDYWNDMLMV ILIKESMKGL X--------- ---------- --------- [649]
#GBI00016442                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00016755                     NNSSQSLSGS PVITPDLDSS TNSAHLK--- ---------- --------- [649]
#GBI00016802                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00016918                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00017675                     PDTWNELLYA SVLIKGX--- ---------- ---------- --------- [649]
#GBI00023219                     PDFEGQQFLT RMQNKTLALV GDSLGRQQF- ---------- --------- [649]
#GBI00023280                     PDTWNELLYA SMLMKGQGMW ASX------- ---------- --------- [649]
#GBI00024264                     PDVWNNLLYA LIMX------ ---------- ---------- --------- [649]
#GBI00024384                     PDIWNELLYS LLLPNHFSLX ---------- ---------- --------- [649]
#GBI00024805                     PDTWNEILFA QLVSPQHKNG SAVIX----- ---------- --------- [649]
#GBI00025061                     ---WSSILAI GGLTRSPTKW X--------- ---------- --------- [649]
#GBI00025062                     ---------- ---------- ---------- ---------- --------- [649]
#GBI00025935                     VDTWNEFVFE IVRRHLNGIM NDMX----- ---------- --------- [649]
#GBI00027480                     PDIWNELLYA ---------- ---------- ---------- --------- [649]
#GBI00029629                     PDVWNHFFFA FIX------- ---------- ---------- --------- [649]
#GBI00029886                     PDTWNELLYY ALLFS----- ---------- ---------- --------- [649]
#GBI00029995                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00004981                     PDTWNELLFA QLLHPGTKST ADKYGHNRSX ---------- --------- [649]
#GMO00005253                     PDVWNHLIYS YLFSSWNGX- ---------- ---------- --------- [649]
#GMO00006383                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00006756                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00006757                     IDTWNEFILE MLKREFQX-- ---------- ---------- --------- [649]
#GMO00007071                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00007072                     PDTWNELLFA SLFMKGNDSX ---------- ---------- --------- [649]
#GMO00007527                     ---------E YLSQEER--- ---------- ---------- --------- [649]
#GMO00011709                     TDYWNDLLMV ILLE------ ---------- ---------- --------- [649]
#GMO00013619                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00015254                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00017486                     PDTWVDVRAA IIMEHMDSKD KCNLYANAMX ---------- --------- [649]
#GMO00019489                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00020816                     ---------- ---------- ---------- ---------- --------- [649]
#GMO00031290                     PDIWNELLYT IIDSX----- ---------- ---------- --------- [649]
#GMO00031544                     PDIWNELLYY DLIFX----- ---------- ---------- --------- [649]
#OSO1G11810                      PDTWNVLFYA ALTGQHX--- ---------- ---------- --------- [649]
#OSO1G30970                      ---------- ---------- ---------- ---------- --------- [649]
#OSO1G42810                      PDTWNLLLLA SLARSPVNVH X--------- ---------- --------- [649]
#OSO1G46350                      PDAWNQLLYA SLAX------ ---------- ---------- --------- [649]
#OSO1G46400                      IDIWNDMLFQ MLLVX----- ---------- ---------- --------- [649]
#OSO1G46410                      VDAWNDMLLH MLSDX----- ---------- ---------- --------- [649]
#OSO1G61460                      PDIWNLILYT RIMSRPQLVX ---------- ---------- --------- [649]
#OSO1G65800                      PDAWNELVIS HLLTNGWRNL AGX------- ---------- --------- [649]
#OSO1G68630                      VDTWNEMLLE TIRREFEGVR SX-------- ---------- --------- [649]
#OSO2G02580                      PDTWVDILAA RILHHFKQAN GX-------- ---------- --------- [649]
#OSO2G49650                      IDLWSELLFH MLVDQX---- ---------- ---------- --------- [649]
#OSO2G49660                      VDAWNELLLH LLRGX----- ---------- ---------- --------- [649]
#OSO2G53380                      PDSWNELLYA QIMVKQHQML HQX------- ---------- --------- [649]
#OSO3G18110                      PDTWNQLFYG LFYQX----- ---------- ---------- --------- [649]
#OSO3G18120                      PDTWNRLLYA HLX------ ---------- ---------- --------- [649]
#OSO3G18140                      PDTWNQFLYA RIASAPWSSD QX-------- ---------- --------- [649]
#OSO3G19470                      PDTWNELLFY HLVSSQQKGV TSX------- ---------- --------- [649]
#OSO3G60300                      PDVWNQLLYA HIVSX----- ---------- ---------- --------- [649]
#OSO3G60340                      PDAWNELLYW KLFFPARDEA IX-------- ---------- --------- [649]
#OSO3G60350                      QDTWNELLYS KLFFPX---- ---------- ---------- --------- [649]
#OSO4G39130                      IDMWNELLLA VLNKYKNSLE DQX------- ---------- --------- [649]
#OSO4G42900                      PDSWNELLYA SLLQAGRGSW RLX------- ---------- --------- [649]
#OSO5G12380                      IDAWNDLVMK MVLGX----- ---------- ---------- --------- [649]
#OSO5G28630                      PDTWNEFIYG HIVSSPQRRP VEPIENQPQR X--------- ---------- [649]
#OSO5G28830                      QDTWNELLYT KLFYPX---- ---------- ---------- --------- [649]
#OSO5G35190                      PDAWNELVFS YLLTNGWRKM ARX------- ---------- --------- [649]
#OSO5G39350                      PDVWNLMLYA RILSRPPAAA GHVAX----- ---------- --------- [649]
#OSO5G51020                      PDTWNLLLFA SWTSLRSX-- ---------- ---------- --------- [649]
#OSO6G10560                      PDSWNELLYA QLLIKQHQML QQX------- ---------- --------- [649]
#OSO6G12800                      INMWNDVLLR RLAELSPPSD ARX------- ---------- --------- [649]
#OSO6G12810                      IDMWNDVLLH KIAETASPPA TNLRX----- ---------- --------- [649]
#OSO6G12820                      VDMWNEMLLQ RLTEISTLDQ DASIFEAPX- ---------- --------- [649]
#OSO6G12840                      VDTFNEILLQ ILSQRX---- ---------- ---------- --------- [649]
#OSO6G12850                      VDTFNEILLQ LLIKKKQQX- ---------- ---------- --------- [649]
#OSO6G12860                      VDTFNEILLQ LLISKRRX-- ---------- ---------- --------- [649]
#OSO6G12870                      IDTFNEILLQ TITRX----- ---------- ---------- --------- [649]
#OSO6G16150                      MDACNDMLFH MLIGX----- ---------- ---------- --------- [649]
#OSO6G16160                      MDACNDMLFH MLIGX----- ---------- ---------- --------- [649]
#OSO6G16170                      IDVWNEMLLQ MLLHQPGAV SVX------- ---------- --------- [649]
#OSO6G16200                      IDVWNEMLLQ MLLHHRSGE FSVX------ ---------- --------- [649]
#OSO6G16240                      VDVWNEMLLQ MLLRHQSX-- ---------- ---------- --------- [649]
#OSO6G16250                      IDVWNNLLLQ IIASSX---- ---------- ---------- --------- [649]
#OSO6G33330                      IDAWNDIIMQ MLAINX---- ---------- ---------- --------- [649]
#OSO6G34060                      PDAWNHIMYA MLLTQQRTX- ---------- ---------- --------- [649]
#OSO6G44900                      PDAWNELIYA SLVLEPKPRS WKHHX----- ---------- --------- [649]
#OSO7G15230                      VDTFNEILVQ LLKRGRX--- ---------- ---------- --------- [649]
#OSO7G15320                      VDTFNEILVQ LLKKRRX--- ---------- ---------- --------- [649]
#OSO7G31550                      TDTWNDLLAA NLLATESX-- ---------- ---------- --------- [649]
#OSO7G46230                      PDAWNELLYA MILKRFSX-- ---------- ---------- --------- [649]
#OSO7G49280                      PDTWNHFLYA HIVAHAAX-- ---------- ---------- --------- [649]
#OSO9G20850                      IDTWNEILQ LVKRWRDSSK X--------- ---------- --------- [649]
#OSO9G23360                      PDMWNELLFA QIX------- ---------- ---------- --------- [649]
#OS11G01370                      PDVWNQILYT RILSKSSPPS PHPPLPPQX- ---------- --------- [649]
#OS11G01570                      PDVWNEFLYA YLMHKX---- ---------- ---------- --------- [649]
#OS11G05080                      PDTWNELVYN LLLKRQSMIG QNVPLVGTKT LKAGWRKLNK YNLTIX--- [649]
#OS11G13800                      PDTWNEILAA QLX------- ---------- ---------- --------- [649]
#OS11G37630                      MDACNDMLFH MLIGX----- ---------- ---------- --------- [649]
#OS12G01380                      PDVWNQILYT RILSKSSPPS THPSLPPQX- ---------- --------- [649]
#OS12G01560                      PDVWNEFLYA YIMHKX---- ---------- ---------- --------- [649]
#OS12G05080                      PDAWNELLYA LFLRRKMVMP HNVSSVGAKR LNTGX----- ---------- [649]
#OS12G33190                      PDTWNEILAA QLX------- ---------- ---------- --------- [649]
#OS12G36750                      PDTWNEILAA QLX------- ---------- ---------- --------- [649]
#PGL00000119                     PDVWNHLLSA LIHX------ ---------- ---------- --------- [649]
#PGL00002746                     ---------- ---------- ---------- ---------- --------- [649]
#PGL00002747                     PDIWNELLYA SLVLEGYIHX ---------- ---------- --------- [649]
#PGL00002880                     PDTWNELLFA RLLFPQPKKP SAQIX----- ---------- --------- [649]
#PGL00004010                     PDMWNELLFA SLLMKGQGVW GTDQX----- ---------- --------- [649]
#PGL00006230                     ---------- ---------- ---------- ---------- --------- [649]
#PGL00006231                     PDTWNELLYA SLILKGFGGX ---------- ---------- --------- [649]
#PGL00008023                     PDTWNELLYG YLLKGYAKD RSTGVSKGX- ---------- --------- [649]
#PGL00008026                     PDIWNELLYG ILVMEGATSP X--------- ---------- --------- [649]
#PGL00008944                     PDVWNNIFYA LIIX------ ---------- ---------- --------- [649]
#PGL00009199                     IDTWNELLLQ MIKDINHFSS X--------- ---------- --------- [649]
#PGL00009692                     ---------- ---------- ---------- ---------- --------- [649]
#PGL00012869                     PDSWNELLYA LLVLRKKAIP GSX------- ---------- --------- [649]
#PGL00013631                     PDTWNELLYA SMLMRAKGVE ETX------- ---------- --------- [649]
#PGL00013729                     ---------- ---------- ---------- ---------- --------- [649]
#PGL00013730                     ---------- ---------- ---------- ---------- --------- [649]
#PGL00014185                     PDTWNHLYG YLLASGYATN RSTGVSRGX- ---------- --------- [649]
#PGL00014800                     PDTWNELLYA SLILKGFGGX ---------- ---------- --------- [649]
#PGL00015011                     PDTWNELLYS ALFFX----- ---------- ---------- --------- [649]
#PGL00015470                     ---------- ---------- ---------- ---------- --------- [649]
#PGL00016142                     IDAWNEFLLE MLKRESMX-- ---------- ---------- --------- [649]
#PGL00016223                     ---------- ---------- ---------- ---------- --------- [649]
#PGL00016224                     PDTWADVLSA LIIEHLNAKT QCLSSSDKRX ---------- --------- [649]
```

```
#PGL00017505                PDIWNELLYS ALFYVX---- ---------- ---------- --------- [649]
#PGL00017520                ---------- ---------- ---------- ---------- --------- [649]
#PGL00017521                ---------- ---------- ---------- ---------- --------- [649]
#PGL00018170                PDTWNELLYA FLIARGFGDX ---------- ---------- --------- [649]
#PGL00018304                IPNYGVRVGV LLSVSSQX-- ---------- ---------- --------- [649]
#PGL00019845                VDTWNEFVFE ITRRHLNGSS NDMX------ ---------- --------- [649]
#PGL00020147                TDYWNDLFMN TLLKERIKVS WX-------- ---------- --------- [649]
#PGL00020630                ---------- ---------- ---------- ---------- --------- [649]
#PGL00021186                PDTWNELLFA QLLFPQPKNL AAQTX----- ---------- --------- [649]
#PGL00023246                PDTWNELLYV TLLFMGFTKX ---------- ---------- --------- [649]
#PGL00023256                ---------- ---------- ---------- ---------- --------- [649]
#PGL00023257                PDTWNELLYV TLLFMGFTKX ---------- ---------- --------- [649]
#PGL00023661                ---------- ---------- ---------- ---------- --------- [649]
#PGL00023830                PDIWNELLYG SLVTEASATP SPRSPLHX-- ---------- --------- [649]
#PP00309G00410              PDTWNELLYA TLLARGQGVW GAPERX---- ---------- --------- [649]
#PP00326G00130              IDTWNDLLVE SLHDKIHNYX ---------- ---------- --------- [649]
#PP00372G00360              PDIWNEQLYA LFLLRASGRA NSQEIKRLFX ---------- --------- [649]
#PP00390G00110              PDTWNELLFA TLLARGQGVW GQPTGQX--- ---------- --------- [649]
#PP00424G00030              IDTWNEYLFA MLQREFIPX- ---------- ---------- --------- [649]
#PP00431G00060              PDTWSQFVLA SLLARGRNTW AQPVRLCDRA KYX------- ---------- [649]
#PPI00001938                ---------- ---------- ---------- ---------- --------- [649]
#PPI00003925                ---------- ---------- ---------- ---------- --------- [649]
#PPI00006425                PDVWNNFLYA LIIX------ ---------- ---------- --------- [649]
#PPI00007112                -DHWDFRVKG X--------- ---------- ---------- --------- [649]
#PPI00007963                ---------- ---------- ---------- ---------- --------- [649]
#PPI00007964                PDIWNELLYG SLVTEASATP SSRSPPLX-- ---------- --------- [649]
#PPI00008021                IDAWNEFLLE MLKRESTX-- ---------- ---------- --------- [649]
#PPI00008416                ---------- ---------- ---------- ---------- --------- [649]
#PPI00008417                PDTWNHLLYA YLLANGYATD RSTGVSX--- ---------- --------- [649]
#PPI00008688                VDTWNEFVFE ITRRHLNGSS NDMX------ ---------- --------- [649]
#PPI00008837                ---------- ---------- ---------- ---------- --------- [649]
#PPI00008838                PDTWNELLYY SLLFKGFGSG X--------- ---------- --------- [649]
#PPI00008849                -DIISTHLNI WLLWX----- ---------- ---------- --------- [649]
#PPI00013412                PDIWNELLYG SLVIEGSVIP SPRYYFHX-- ---------- --------- [649]
#PPI00013432                IDSWSEILLE ILQHEX---- ---------- ---------- --------- [649]
#PPI00014229                PDIWNELLYG SLVLEGGTSP X--------- ---------- --------- [649]
#PPI00014345                PDTWVDVLSA VIVEHLNAKR QCLLSLDKX- ---------- --------- [649]
#PPI00015461                PDTWNELLYD SLLLKGFGGX ---------- ---------- --------- [649]
#PPI00015845                LKIFSDTISQ TFX------- ---------- ---------- --------- [649]
#PPI00016354                ---------- ---------- ---------- ---------- --------- [649]
#PPI00016634                PDAWNELVFA SLVMEEKWKH RX-------- ---------- --------- [649]
#PPI00016712                PDTWNELLYA SLLFKGFGVX ---------- ---------- --------- [649]
#PPI00029676                ---------- ---------- ---------- ---------- --------- [649]
#PPI00029965                PDMWNELLFG YLLMEGRGV- ---------- ---------- --------- [649]
#PPI00029966                X--------- ---------- ---------- ---------- --------- [649]
#PPI00033679                PE-------- ---------- ---------- ---------- --------- [649]
#PPI00039894                PDTWNELLYS ALFFX----- ---------- ---------- --------- [649]
#PPI00044132                PDIWNELLYG SLVTEASATP SSRSPPLX-- ---------- --------- [649]
#PPI00057730                PEEEESLKT- ---------- ---------- ---------- --------- [649]
#PPI00057838                ---------- ---------- ---------- ---------- --------- [649]
#PPI00058297                X--------- ---------- ---------- ---------- --------- [649]
#PPI00058298                PDMWNELLFG YLLMEGRGVW GTDIMLX--- ---------- --------- [649]
#PPI00058512                IDTWNEILLQ MIQDIX---- ---------- ---------- --------- [649]
#PPI00063580                PDKDYEIGGG NLX------- ---------- ---------- --------- [649]
#PPI00063597                PDTWNELLYA SMVIRGKGLR TX-------- ---------- --------- [649]
#PPI00063681                ---------- ---------- ---------- ---------- --------- [649]
#PPI00063682                ---------- ---------- ---------- ---------- --------- [649]
#PPI00064161                ---------- ---------- ---------- ---------- --------- [649]
#PPI00065500                PDTWNHLLYG YLLANGYETD RSKX------ ---------- --------- [649]
#PPI00065764                PDTWNELLYA FLIAKGFGDX ---------- ---------- --------- [649]
#PPI00069149                ---------- ---------- ---------- ---------- --------- [649]
#PPI00069197                ---------- ---------- ---------- ---------- --------- [649]
#PPI00069990                PDSWNELLYA FLVLRKKAIP GSX------- ---------- --------- [649]
#PPI00069991                ---------- ---------- ---------- ---------- --------- [649]
#PPI00070464                ---------- ---------- ---------- ---------- --------- [649]
#PPI00071160                PDVWNHLLSP LIHX------ ---------- ---------- --------- [649]
#PPI00072989                PDIWNELLYS ALFYPX---- ---------- ---------- --------- [649]
#TBA00001677                ---------- ---------- ---------- ---------- --------- [649]
#TBA00004346                PDSWNELFYA SLIFGKHX-- ---------- ---------- --------- [649]
#TBA00006917                PDTWNQLLYS SLLFX----- ---------- ---------- --------- [649]
#TBA00009852                PDNWNELLSS YLLX------ ---------- ---------- --------- [649]
#TBA00010198                ---------- ---------- ---------- ---------- --------- [649]
#TBA00011032                KPSX------ ---------- ---------- ---------- --------- [649]
#TBA00011296                PDQWNHLLYG YLIASGYGEL GTRDSRRX-- ---------- --------- [649]
#TBA00011691                ---------- ---------- ---------- ---------- --------- [649]
#TBA00012144                CDTGNENPCY VHEHHKNDRL X--------- ---------- --------- [649]
#TBA00012145                -KTSSPIISC VYIGQTFGDC FQX------- ---------- --------- [649]
#TBA00015462                PDTWNEMLFA SLLMKGWGFQ GTDYX----- ---------- --------- [649]
#TBA00015654                LPAFTETKAC DY-------- ---------- ---------- --------- [649]
#TBA00016428                TDYWNDLLIV TL-------- ---------- ---------- --------- [649]
#TBA00017489                PDTWNELLYA SLVQPAX--- ---------- ---------- --------- [649]
#TBA00017684                ---------- ---------- ---------- ---------- --------- [649]
#TBA00017880                ---------- ---------- ---------- ---------- --------- [649]
#TBA00017881                PDTWNELLYA TLIARGYGDX ---------- ---------- --------- [649]
#TBA00019818                ---------- ---------- ---------- ---------- --------- [649]
#TBA00019847                PDAWNELIYS SLVMEKRANH NX-------- ---------- --------- [649]
#TBA00021216                PDTWNELLFA QLFFPQGKNL SVSMVSX--- ---------- --------- [649]
#TBA00021480                VDTWSEFLLE LLKRDSEX-- ---------- ---------- --------- [649]
#TBA00024636                ---------- ---------- ---------- ---------- --------- [649]
#TBA00025353                ----EQMCYG ---------- ---------- ---------- --------- [649]
#TBA00026111                ---------- ---------- ---------- ---------- --------- [649]
#TBA00026417                ---------- ---------- ---------- ---------- --------- [649]
#TBA00027512                VDTWNELVFE IVRRNLNGID AX-------- ---------- --------- [649]

#PGL00026077                PDTWNELLFA QLLFPQPKKR STQIX----- ---------- --------- [649]
#PGL00026078                RPIYSGKMFK MX-------- ---------- ---------- --------- [649]
#PGL00028109                PDTWNELLYA SLLFKGFGGG X--------- ---------- --------- [649]
#PP00001G00240              IDTWNDLLVE SLRDIIFKX- ---------- ---------- --------- [649]
#PP00002G00270              IDTWNDMLVH SLRDVIFKX- ---------- ---------- --------- [649]
#PP00004G00340              PDTWNELLYA SLLARGQGVW GEPIRYX--- ---------- --------- [649]
#PP00009G00100              PDTWNEILFA DLATKLTSTA IASRAEMRX- ---------- --------- [649]
#PP00023G00570              IDTWNDLLVQ SLQDVIHRX- ---------- ---------- --------- [649]
#PP00057G00940              PDTWNELLYA ALLARGQGVW GKFVX----- ---------- --------- [649]
#PP00101G01110              PDTWNEILFA DLARKFKSRE SSTVAX---- ---------- --------- [649]
#PP00106G00770              IDTWNDLLVE SLRDAVFGSX ---------- ---------- --------- [649]
#PP00154G00820              IDTWNDLLVQ LLQDVIHRX- ---------- ---------- --------- [649]
#PP00242G00420              IDTWNEMLME TLQQQMQPX- ---------- ---------- --------- [649]
#PP00297G00240              PDTWVDMLVD VVLKHVQSAQ X--------- ---------- --------- [649]
```

Figure S5: ML phylogenetic tree of ESK1 homologs in several land plants. (continued). In blue is highlighted AtESK1, and in red the ESK1 closest homologues clade.


PPI00014345
PGL00016223
GBI00009760
AT5G64470 TBL12
Eucgr.I02565
ATR 00040G01910
PP00297G00240
BD3G01660
OS02G02580
CMI00008379
TBA00026111
PGL00016224
GBI00003825
GMO00017486
GBI00007382
PPI00015845
GBI00015156
PGL00020147
TBA00016428
CMI00011961
ATR 00057G01830
BD1G26890
OS07G31550
Eucgr.G01129
AT2G14530 TBL13
GMO00011709
TBA00011691
CMI00013875
GBI00007295
TBA00027512
CMI00027327
GBI00025935
GMO00006383
PP00424G00030
PGL00019845
PPI00008688
ATR 00025G02240
AT4G25360 TBL18
AT5G51640 TBL17/YLS7
BD2G58580
BD2G58570
OS01G68630
TBA00024636
ATR 00088G00780
BD1G44920
OS06G12820
OS06G12800
OS06G12810
AT3G28150 TBL22/AXY4-L
OS04G39130
OS06G16170
BD1G43730
OS06G16200
OS06G16240
BD1G43740
OS11G37630
OS06G16150
OS06G16160
BD3G56217
OS02G49650
BD3G56210
OS02G49660
BD1G43717
OS06G16250
BD2G45350
OS01G46400
BD2G45360
OS01G46410
AT5G15890 TBL21
AT3G02440 TBL20
AT5G15900 TBL19
ATR 00088G00740
AT1G70230 TBL27/AXY4
OS09G20850
BD1G44890
OS06G12870
BD1G44900
OS06G12850
OS06G12860
OS06G12840
OS07G15230
OS07G15320
GMO00006756
CMI00009108
GBI00016802
PPI00008021
PGL00016142
GMO00006757
CMI00009353
CMI00019649
GBI00006281
TBA00021480
ATR 00093G00040
BD2G32220
OS05G12380
BD1G38400
OS06G33330
AT4G11090 TBL23
AT4G23790 TBL24
AT1G01430 TBL25
AT4G01080 TBL26
PGL00013729
PPI00013432
GBI00016918
PGL00013730
CMI00003325
PP00242G00420
PP00001G00240
PP00002G00270
PP00023G00570
PP00154G00820
PP00106G00770
PP00326G00130
TBA00019818
CMI00018183
PPI00001938
CMI00012342
PGL00009199
PPI00058512
0.50

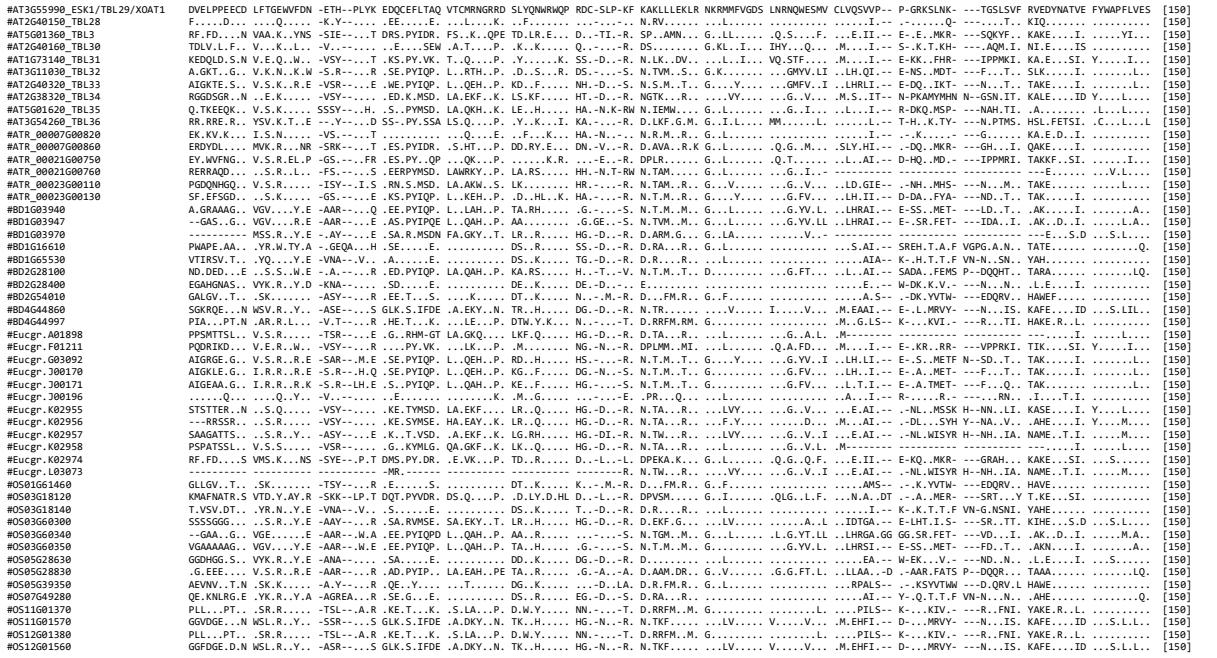

GMO00007527
TBA00009852
GMO00013619
PPI00069991
PGL00012869
PPI00069990
GBI00002300
GBI00029995
TBA00010198
ATR 00045G02000
AT3G62390 TBL6
Eucgr.K03547
BD1G64470
OS03G19470
TBA00019847
PPI00016634
PPI00057838
CMI00014316
GBI00002329
ATR 00019G03720
BD1G31440
OS06G44900
AT5G20590 TBL5
Eucgr.C00073
GBI00013823
GBI00013099
GMO00020816
ATR 00085G01050
AT1G60790 TBL2
Eucgr.H02546
BD5G15190
OS04G42900
TBA00012144
GBI00023280
PPI00063597
GBI00002328
PGL00013631
GBI00004587
CMI00027594
GMO00019489
BD1G46280
OS06G10560
BD3G57480
OS02G53380
Eucgr.B03636
Eucgr.A02463
AT3G12060 TBL1
AT5G06700 TBR
AT5G49340 TBL4
ATR 00048G01150
ATR 00040G02130
PGL00000119
PPI00071160
GBI00025062
GBI00025061
GBI00024264
TBA00026417
PPI00006425
PGL00009692
PGL00008944
CMI00007639
GBI00024384
CMI00016461
GMO00015254
PGL00015470
PGL00008026
PPI00014229
PPI00007963
CMI00001152
PPI00013412
PPI00044132
PPI00007964
PGL00023830
PGL00002747
GBI00027480
Eucgr.J00196
AT2G40150 TBL28
AT3G55990 TBL29/XOAT1/ESKIMO1
BD4G44997
OS11G01370
OS12G01380
OS05G39350
BD2G54010
OS01G61460
BD1G65530
OS03G18140
BD1G16610
OS07G49280
OS05G28630
BD2G28400
AT2G40160 TBL30
ATR 00021G00750
AT1G73140 TBL31
Eucgr.F01211
OS03G18120
AT5G01360 TBL3
Eucgr.K02974
BD2G28100
OS05G28830
BD1G03947
OS03G60340
OS03G60350
BD1G03940
AT3G11030 TBL32
Eucgr.G03092
AT2G40320 TBL33
Eucgr.J00170
Eucgr.J00171
ATR 00023G00130
BD1G03970
OS03G60300
AT2G38320 TBL34
Eucgr.K02956
Eucgr.K02955
Eucgr.K02957
Eucgr.L03073
Eucgr.A01898
0.50


Eucgr.J00196
AT2G40150 TBL28
AT3G55990 TBL29/XOAT1/ESKIMO1
BD4G44997
OS11G01370
OS12G01380
OS05G39350
BD2G54010
OS01G61460
BD1G65530
OS03G18140
BD1G16610
OS07G49280
OS05G28630
BD2G28400
AT2G40160 TBL30
ATR 00021G00750
AT1G73140 TBL31
Eucgr.F01211
OS03G18120
AT5G01360 TBL3
Eucgr.K02974
BD2G28100
OS05G28830
BD1G03947
OS03G60340
OS03G60350
BD1G03940
AT3G11030 TBL32
Eucgr.G03092
AT2G40320 TBL33
Eucgr.J00170
Eucgr.J00171
ATR 00023G00130
BD1G03970
OS03G60300
AT2G38320 TBL34
Eucgr.K02956
Eucgr.K02955
Eucgr.K02957
Eucgr.L03073
Eucgr.A01898
Eucgr.K02958
BD4G44860
OS11G01570
OS12G01560
ATR 00023G00110
ATR 00023G00100
AT5G01620 TBL35
OS05G51020
BD2G43447
OS01G42810
Eucgr.A01903
ATR 00021G00630
ATR 00021G00670
CMI00015101
PPI00070464
PPI00003925
PGL00023246
PGL00023256
PGL00028109
PPI00016712
PPI00008837
PPI00016354
PGL00006231
PGL00006230
PGL00014800
PPI00015461
PGL00023257
PPI00008838
TBA00017684
TBA00017489
TBA00004346
PPI00069149
CMI00000518
PGL00018304
GBI00017675
ATR 00114G00150
AT3G54260 TBL36
Eucgr.J00985
TBA00025353
GBI00016442
CMI00014513
PGL00015011
PPI00039894
TBA00006917
GBI00029886
GMO00031544
GMO00031290
PPI00063682
PPI00072989
PPI00063681
PGL00017505
PGL00017520
PGL00017521
BD1G65550
OS03G18110
AT2G30010 TBL45
AT5G58600 TBL44/PMR5
Eucgr.B02240
Eucgr.J01893
ATR 00021G00760
ATR 00021G00710
ATR 00021G00730
ATR 00021G00700
ATR 00021G00680
ATR 00021G00690
Eucgr.F03235
AT3G14850 TBL41
Eucgr.F01363
AT1G78710 TBL42
Eucgr.F01364
AT2G30900 TBL43
Eucgr.A01950
Eucgr.K01183
Eucgr.K01184
Eucgr.E02713
AT1G29050 TBL38
AT2G34070 TBL37
Eucgr.A01900
Eucgr.A01902
AT2G31110 TBL40
AT2G42570 TBL39
BD2G45320
OS01G46350
BD2G07000
OS01G11810
BD1G32150
OS06G34060
0.50

Figure S6: Alignment of ESK1 closest homologs in five angiosperm species. This alignment was used to build the Maximum likelihood phylogenetic tree in Chapter 6. The variable N-terminal region was removed from the original alignment.

Key:

AT *Arabidopsis thaliana*
ATR *Amborella trichopoda*
BD *Brachypodium distachyon*
Eucgr. *Eucalyptus grandis*
OS *Oryza sativa*



```
#AT3G55990_ESK1/TBL29/XOAT1   GIKCALETTP I--LNMSM-P ---FSVGTDY ---RLFSVAE NVTHSLN--- --VPVYFLNI TKLSEYRKDA HTSVHTIRQG KMLTPEQQA- DPNTYADCIH WCLPGLPDTW NEFLYTRIIS RS-------- ---------- - [441]
#AT2G40150_TBL28              ..R..... .LNMSFNVAY GQFSA..... ---...P... ...Q..K--- --...IH.... .A........ ....Y..K.. .L..R...N- ..ANF..... .......... ......H... .R-------- ---------- - [441]
#AT5G01360_TBL3               .T..FN..K. .--KDKK--F ---WGT.SNK ---QMMK.VS S.IKHMT--- --TH.TVI.. .Q.....I.. ....Y.ETG. .I..A..R.- ..MHH..... .......... .RI.LAHL-- ---------- ---------- - [441]
#AT2G40160_TBL30              .S..EK..E. .---....K-. ---IN...NR ---..YEI.L .A.K.T--K- ---..IH.... .TM......G ...FYGSIN. .LM....KL- ..R.F...Y. ........S. ...L.SLFLLH .ERQKVTGRD ERATAK----- ---------- - [441]
#AT1G73140_TBL31              DGT.YD.LY. .--DKR.--Y ---WGT.SNQ ---EIMKIVG D.LSRVG--- --EN.T.... .Q.......G ..T.YGE.R. .L..K..R.- ..KN.G..... .....V.... ...I..AYLLR SHRNFF---- ---------- - [441]
#AT3G11030_TBL32              .KN.YNQ... .--QD.N--H ---WPSDCSK ---T.MK.IG EELDQRAE-- --F..TV... .Q..G..... ...IYKKQ-W SP..K..L.- N.AS.S.... ......Q... ..LFFAKLFY P--------- ---------- - [441]
#AT2G40320_TBL33              .QN.YNQ..L .--EDP.--Y ---WGSDCRK ---SIMK.IG E.FGRSK--- --T.ITL... .QM.N..... ...IYKKQ-W SP..A..LE- N.TS....V. ......Q... ..L.FAKLFY T--------- ---------- - [441]
#AT2G38320_TBL34              NQN.YG.ASL .--DKEG--Y ---TGR.S.P ---KMMR.L. ..LDG.KNRG --LNMQMI.. .Q......EG .P.IYKKQ-W GTVKENEIS- N.SSN..... .....V..V. ..L..AY.LD HHSS------ ---------- - [441]
#AT5G01620_TBL35              EGN.YG.KK. .--EEE.--Y ---WGS.S.I ---PTMRMVK R.LER.G--- --PK.SVI.. .Q.......G .P..YRKF-W EP.NEDRLK- N.AS.S..T. ..V..V..V. .QL.FHFL-- ---------- ---------- - [441]
#AT3GS4260_TBL36              .QM.YNQKH. LPS.SS.--- ----TKPHVP ---QQSR.LN K.LRTMK--- --YR..LYD. .TM.A..R.G .P..FKRAMH EEEKHHRI.- G.SS--..S. .....V..I. ..M.SSI.LT NAV------ ---------- - [441]
#ATR_00007G00820              .....K..M. .--.D.P--T ---VQ..... ---..V.A..S .I.S.MKK-- --...S.I.V .A...L.... ....V..... ....R..K.- ..A..... .......... .....QL.F .--------- ---------- - [441]
#ATR_00007G00860              ..R.YN..R. V--..KRG--Y ---WGS.S.G ---..M..VG SIVKRMK--- --...SVI.. .Q...H.I.G ....Y.ETH. .V.ND..K.- ..LHN..... ....V.... .QI..SYL-- ---------- ---------- - [441]
#ATR_00021G00750              SGN.YN..H. .--QGS---Y ---LG..SST ---EMMDIVV .TLR.M.--- --.G.TVIK. .Q........ ...LY.E.A. .L..A..RS- E......... ....V.... ..I...YLVY HK-------- ---------- - [441]
#ATR_00021G00760              ---------- ---------- ---------- ---------- ---------- ---------- -------- ---------- ---------- ---------- ---------- ---------- - [441]
#ATR_00023G00110              .GN.YN..E. .--RDER--Y ---WGE.S.K ---.MIDAMD Q.IAGMREKG --.S.S.... .Q.......G .P.IYKKQ-W EP.K...L.- ..AS..... .....V.... ..L..ASLF. HQ-------- ---------- - [441]
#ATR_00023G00130              EGN.YN.... .--EDPT--Y ---WGR.SSK ---GIMR.LG Q.LSGMK--- --...TL... .Q........ ...IYKKQ-W NP.....L.- N.VS....V. ......Q... ..L..AKLFF P--------- ---------- - [441]
#BD1G03940                    DGN.YNQ... .--RDL.--Y ---WGP..SK ---G.MR.IG E.FSTSK--- --...GIV.. .Q........ ..QIYKKQ-W NP.....I.- N.KS....T. ......Q... ..L..SKLFF P--------- ---------- - [441]
#BD1G03947                    .ND.YGQK.. .--TGDE--- ---HASY.SR ---AMLK..D E.LGGSRSRR AV...GVV.V .RM....R.. ..Q.YREQW. .TGVAGK.E- --QSD...T. .....V..A. ..L..WKLFF PAADDQAL-- ---------- - [441]
#BD1G03970                    YG..YN..E. .--ATDAR-Y ---VGGAGRR GYAA.AGAV. DAVRG.GARG --.G.RL.DV .RI..R.... .P...RRQ-W EP..EA.RRE ..AG....V. .....V..V. .HL..AH.VA ---------- ---------- - [441]
#BD1G16610                    AV...M..E. .--M.RT--W ---VNI...W ---..HG..Q G.LG.MRR-- --...HLVD. .A...F.... ......L... .L..R....- ..RA..... .......... .H...A...A D.RSSSSTTP ASSLDISATS S [441]
#BD1G65530                    .....M..M. .--K.RTA-S ---LD....W ----..YAG.Q D.LQTFR-R- --...H.VD. .A...L.... ......L... .L........- ..K..... .......... .....A..V. SPSTTTEQ-- ---------- - [441]
#BD2G28100                    .GN.YN..AM V--.QDAE--Y ---WGTDSRR ---SVMR.IG DILDGDGAD- --..LT.... .Q..M..... ...IYKKQ-W NP..A..L.- ..KS....V. ......Q... ..L...KLFY P--------- ---------- - [441]
#BD2G28400                    N..FS..Q. A--I.YTK-E ---LEL...W ---D...ASH GF..KAMK-K- ----..H.I.. .A...I.... ......L... .L..K.....- N.RKF..... .......... ...I.GH.V. SPLQ---QQI EDQSQR---- - [441]
#BD2G54010                    ..R..K..M. L--..WRGPI ---W-L...W ---DM.K..S ...GAAAPR- --..IT.VD. .TM..R...G .......... QV.......- ..G..... .....V..I. .SL....M. .PETGIATTR ---------- - [441]
#BD4G44860                    KN..LN.SE. .--VKPG--Y ---KGAT... ---SMMEK.Q EIFRP.EEKG --IH.QV..F .Q..DH.I.. .PT.FRRQ-F VP..AA.V.- ..SS....T. .....V..V. .S...SYLVH K--------- ---------- - [441]
#BD4G44997                    A...FN..L. V--..YTE-. ----.DLNH.M ---..MYDLVA STSR.MK-K- --...TLID. .RM.D..... ....Y...RD RL...K.R.- ..EKF..... .....V..V. .TV.F...F. K.SPP-PLPS LPPQ------ - [441]
#Eucgr.A01898                 .HN.FG..E. .--TKEG--Y ---WGS.SNR ---EMMR.LQ KALNK.SKRG --LN.QM... .Q........ .V.IYKGQ-A .N..QH.I.- N.A....F. ........V. ..L..AY.LQ N--------- ---------- - [441]
#Eucgr.F01211                 DGN.FG.SY. .--QGP---Y ---WGS.SNL ---EIMKIVR HALRQ.K--- --IN.TL... .QM..F.... ....F.E.K. .L..KA.RS- ..K.FG.... .....V..A. ..I..AYLLK DHRTIL---- ---------- - [441]
#Eucgr.G03092                 .GN.YNQ... .--EDPN--Y ---WGSDSRK ---SIMQ.IG E.FSKSK--- --...IT.... .Q..N..... ..MIVKKQ-W SP.....L.- N.VS....V. ......Q... ..L.FAKLFY P--------- ---------- - [441]
#Eucgr.J00170                 .GN.YNQ..I .--EDPN--Y ---WGSDCRK ---GVMR.IG E.FSKTK--- --...IVL... .Q..S..R.. ...IYKKQ-W HP..A..V.- N.AS....V. ......Q... ..L.F.KLFY T--------- ---------- - [441]
#Eucgr.J00171                 .GN.YN..AM .--EDPN--Y ---WGSDCSK ---SIMR.IG Q.FSKAE--- --...ITV... .Q..S..... ...I.KKQ-W NP.....L.- N.VS....V. .....H... ..L.F..LFY P--------- ---------- - [441]
#Eucgr.J00196                 .....K..M. V--...TTA-Q ---LN....R ----...VI.D ...R.M--K- --L..H.M.. .T........ ........ ........- ..Q..... .......... ...I....LQ H.-------- ---------- - [441]
#Eucgr.K02955                 .QN.YA..E. .--AKEG--Y ---WGS.S.P ---GMMR.V. S.IDQ.KARG --LG.QI... .Q......E. .PTI.RRQ-W EP..E..K.- K.TS....F. .....V..V. ..L..AH.FQ NRTQRLENRV EGL------- ---------- - [441]
#Eucgr.K02956                 .RN.LG..EQ .--TREG--Y ---WGN.SVP ---AMMR.V. S.IEK.KTRG --.TIHV... .Q......E. .PAIYRKQ-W HP..K..L.- Y.IS....L. .....V..V. ..L...Y.FR YIHT------ ---------- - [441]
#Eucgr.K02957                 .HN.YQ..E. .--TREG--Y ---SGS.SRP ---E.MRAV. A.IGR.RARG --LG.RI... .Q........ .P.IYKRQ-W HP.SR..L.- ..TS....F. .....VS.V. ..L..AY.LR NRSPC----- ---------- - [441]
#Eucgr.K02958                 DHN.FE..E. .--TEGG--Y ---RVRRSTR ---GMMHTLR KAVNK.QKRG --LN.HV.D. .......... .S.IYKEQ-T .N..QH.M.- N.V....F. ........V. ..L..AY.LH N--------- ---------- - [441]
#Eucgr.K02974                 ....FN..K. V--MKKN--H ---WGT.S.K ---.IMK.VS S.VEKMK--- --...T.I.V .Q.....I.. .S.IY.ETG. SL..D..K.- ..LHH..... .....V.... .QI..AHL-- ---------- - [441]
#Eucgr.L03073                 .HN.YQ..E. .--TREG--Y ---SGS.SRP ---E.MRAV. A.IGR.RARG --LG.RI... .Q........ .P.IYKRQ-W HP.SR..L.- ..TS....F. .....VS.V. ..L..AY.LR NRSPC----- ---------- - [441]
#OS01G61460                   ..R..K..M. L--..WHGPI ---W-L...W ---DM.HA.A ..SRTAATR- --..IT.VDV .TM..R...G .......... .V........- ..G..... .....V..I. .LI....M. .PQLV----- ---------- - [441]
#OS03G18120                   ..R.YN..W. .--TQRG--Y ---RGS.S.R ---..MME.MS D.LGRMR--- --T..TL... .Q.T.H.V.. .V..V.ETG. LLV.D.EKT- ..QR.T.... ..I..V.... .RL..AHL-- ---------- - [441]
#OS03G18140                   .....M..L. .--T.RTT-S ---LD....W ----..YAG.Q E.LQTFR-R- ----.HLVD. .A...L.... ......L... .L......S- ..K..... .......... .Q...A..A. APWSSDQ--- ---------- - [441]
#OS03G60300                   NHG.YG..E. .AAEEYR--- ---GTS...M ---AFARAV. AEARR.GERS --.A.RLI.V .R...R.... .P...RRY-W DPV.D..RR- N.SS....T. .....V..V. .QL..AH.V. ---------- ---------- - [441]
#OS03G60340                   .GE.YDQ... VGAADAA--S ---YCGS.SR ---.MVQ..G E.LGASR--- --...GVV.V .RM..L.R.. ..Q.YREQRW AKP.A..L.A ..RS....T. .....V..A. ..L..WKLFF PARDEAI--- ---------- - [441]
#OS03G60350                   DGN.YNQ... .--RDL.--Y ---WGP..SK ---G.MR.IG E.FSTSK--- --...GIV.. .Q........ ..QIYKKQ-W NP.....I.- N.KS....T. ......Q... ..L..SKLFF P--------- ---------- - [441]
#OS05G28630                   A...FS..Q. A--I.YTK-K ---LE....W ---D...T.H H..KAMK-R- --...H.I.. .A...I.... ....N.L... .L..K..K.- N.RKF..... .......... ...I.GH.V. SPQRRPVEPI ENQPQR---- - [441]
#OS05G28830                   .GN.YG..EM .--GDAA--Y ---WGSDSRR ---GVMRAIG E.LDGDGAD- --...T...V .Q..L..... ....YKKQ-W TPP....L.- ..K.....V. ......Q... ..L...KLFY P--------- ---------- - [441]
#OS05G39350                   .VR..K.DA. V--Q.WHGPL ---W-L...W ---DM.RA.R .ASRAAGR-- --...T.VDV .AM..L...G .......... RV.......- ..A..... .....V..V. .LM..A..L. .PPAAAGHVA ---------- - [441]
#OS07G49280                   AV...M..Q. .--V.RTSGG ---LDI...W ----..HG..R G.LR.MR-R- --..G.R.VD. .A...L.... ......L... .L........- ..R..... .......... .H...AH.VA HAA------- ---------- - [441]
#OS11G01370                   ....VK..L. V--..YTK-. ---LDLNH.M ---..MYDLVA K.AKNMKN-- --...SLID. .RM.D..... ...LYS.... .L.....K.- ..QK..... .....V..V. .QI....L. K.SPPSPHPP LPPQ------ - [441]
#OS11G01570                   NN..LN..E. .--QIED--Y ---R.AT... ---GMMDK.K EIFGT.EPKG --IH.QI... .Q........ .PTIFRRQ-Y VP..K..I.- N.SI....T. .....V..V. .....AYLMH K--------- ---------- - [441]
#OS12G01380                   ....VK..L. V--..YTK-. ---LDLNH.M ---..MYDLVA K.AKNMKN-- --...SLID. .RM.D..... ...LYS.... .L.....K.- ..QK..... .....V..V. .QI....L. K.SPPSTHPS LPPQ------ - [441]
#OS12G01560                   NN..LN..E. .--QIED--Y ---R.AT... ---GMMDK.K EIFGT.EPKG --IH.QI... .Q........ .PTIFRRQ-Y VP..K..I.- N.SI....T. .....V..V. .....AY.MH K--------- ---------- - [441]
```